EXPERTISE:

AN ANALYSIS OF TRANSFER EFFECTS IN WELL-STRUCTURED PROBLEM SOLVING DOMAINS

Peter Charles Wright
Department of Psychology
University of Edinburgh

Thesis Submitted for the degree of PhD
January 1985

ABSTRACT

This thesis is concerned with expertise in problem solving. An expert solution can be distinguished from simple success by virtue of the fact that such solutions are tempered by considerations of economy and transferability. On this latter point, two characterisations of expertise have emerged in the literature. The first views expertise as the ability to use domain specific knowledge in order to develop highly efficient solutons but solutions applicable to only a narrow class of problems. The second views expertise as the ability to deploy problem solving heuristics which, while now necessarily providing the most efficient solution are applicable to a diverse range of problem solving tasks.

While the psychological literature abounds with studies of problem solving there has been very little work on the transferability of problem solutions, and what little work there is fails to distinguish between these two characterisations of expertise. Recent work in artificial intelligence on the other hand has made a clear distinction between the two characterisations but offeres no empirical evidence for such distinctions.

The experiments reported in this thesis demonstrate that both forms of expertise are observable in the laboratory but that task specific expertise produces by far the most dramatic transfer effects. In attempting to investigate the psychological mechanisms underwriting such expertise, it is argued that a clear distinction must be made between representation and strategy. Expertise consists in the subject being able to identify a problem, or class of problems as solvable by recourse to some general representational resource and being able to deploy task-specific information processing strategies dedicated to interpreting novel problems with respect to this resource.

The functional goals at work in the problem solving domain constrain a subject´s choice of representational space and, even for expert subjects, surface structure variation in the problem domain involves the subject in extra processing operations to make problems congruent with preferred representations. Thus, contrary to some approaches to problem solving, expertise cannot be defined as that which renders formally equivalent problems psychologically equivalent. On the contrary, for expert subjects, surface structure plays an important part in determining the information processing strategies necessary for solution.

Acknowledgements

The work reported in this thesis was funded by SERC grant number 80310005.

I would like to thank my supervisor, Dr. Brendan McGonigle for his patient and unerring guidance, and also Janet Wright for considerable help in the preparation of this document.

DECLARATION

I declare that this thesis has been composed by myself, and that the work presented herein is my own.

FOR MY FATHER

CONTENTS

CHAPTER FIVE:

CHAPTER SIX:

CHAPTER ONE

PSYCHOLOGICAL STUDIES OF PROBLEM SOLVING AND THE CONCEPT OF EXPERTISE

Introduction

This thesis is about problem solving. In particular it is concerned with an analysis of expertise in problem solving. The term problem solving could be be seen to cover a very large area of psychological research, particularly if it is understood broadly as that area of psychology involved in an investigation of goal-directed behaviour. Under this interpretation the term can be used to describe areas as diverse as animal learning, and human inference. But even within this broad area there has been very little research that can be said to concern itself with an understanding of expertise.

Learning theorists for example, have tended to use criteria of ´success´ to indicate when learning has taken place. Animals are said to have learned when they are able to gain reward by making some arbitary response, the precise nature of which is of little concern to the experimenter. All that is required is that the animal be able to succeed in gaining reward by whatever means possible. But expertise can be distinguished from simple success in as much as it presupposes success but is tempered by other considerations. An expert solution will be tempered by considerations of economy for example, providing a solution which involves the subject in the minimum of effort. An expert solution will also be distinguishable by the extent to which it can be used to solve problems other than the one in which the solution was forged.

But while the transferability of a solution is indeed a hallmark of expertise, not all transfer experiments tap an expertise factor. In the area of animal discrimination learning for example, it was common to use a simple two-problem ´train and test´ paradigm in which subjects were trained on just one problem and then tested on a second problem (Reese 1968, Riley 1968). This paradigm has also been used extensively in the human problem solving literature (Wertheimer 1945, Luger 1976, Johnson-Laird 1972). But this kind of paradigm does not guarantee that any level of expertise has been achieved on the first problem, hence it is really only testing the extent to which any solution that the subject cares to generate in response to that first problem is transferable to the second (cf. McGonigle and Jones 1978).

In contrast to the above approach, the learning set paradigm of Harlow used a multiple problem environment, in which subjects solved each problem many times as well as solving many different problems. Thus there were both intra- and inter- problem learning components to ensure that some level of expertise is generated within the testing situation. Using this transfer paradigm, Harlow found that test-sophisticated animals could solve novel problems in a single trial whereas naive subjects take many trials to solve their first problem.

But while the learning set studies can be used to make a distinction between success and expertise in terms of economy and transferability, the actual nature of the expertise generated within the learning set paradigm is still not known. It is possible in principle to distinguish between two characterisations of expertise. In the first instance expertise may be viewed as the acquisition of knowledge about a limited domain of problems which while allowing all problems within that domain to be solved efficiently, nevertheless produces solutions which are confined to that domain. This is the popular conception of expertise, and the one which gives warrant to beliefs that for example a high level proficiency in the game of squash tends to interfere with one's ability to play badminton. But a second view of expertise might be that it is possible to develop some quite general problem solving skills which enable problems of diverse sorts to be tackled with some level of proficiency. So for example we might expect that there are certain skills required by the game of Bridge such as remembering which cards have been played, which will also be of use when playing other card games.

It would seem that in order to distinguish between these two possible forms of expertise it would be necessary not only to consider the transferability of problem solutions but also the particular domain over which solutions are transferable. Expertise of the domain-specific sort would of course only be expected to range over problems of a single class. If a domain of problems could be formally defined in terms of its logical or mathematical structure, we would expect expertise of this sort to render such formally equivalent problems psychologically equivalent. Thus a subject would be able to identify problems as members of the class and to use a solution method that stood as a solution to that whole class of problems (See for example the work of Dienes and Jeeves 1970).

In contrast, general expertise would not be so constrained since, by definition it is not founded in the subject's knowledge of a particular class of problems but involves knowledge of certain problem solving methods which could be applied to a broad range of goal-directed activities (see for example the work of Newell and Simon 1972). Harlow claimed to be studying just such a general problem solving skill, namely learning-to-learn, yet he only examined transfer of learning within the confines of a single class of problems.

These considerations establish a range of descriptive criteria which are necessary for an adequate study of expertise. Such a study would need to look at transfer effects in multiple problem domains in which problems can be classified into problem-types in an a priori way on the basis of some structural analysis, and this analysis must be of psychological relevance. It would be necessary to use this taxonomy of problems in order to differentiate between the two kinds of expertise outlined above and to assess which of the two kinds is observed within the confines of the laboratory situation. Finally it would be necessary to examine the psychological mechanisms that underwrite such expertise. In the sections that follow, research on problem solving is reviewed against these criteria in order to assess what further experimentation is required.

Success Versus Expertise

The early approaches to problem solving from the field of learning research were not concerned with identifying expertise factors in problem solving. Rather they were more straightforwardly concerned with the mechanisms by which the subject could acquire any solution to a problem however unsophisticated.

In Thorndike´s experiments for example, cats were placed in puzzle boxes, escape from which was contingent upon them pressing a lever to open the door. Here the animal was seen to respond haphazardly in a trial and error fashion with apparently no directed problem solving behaviour. The animal would continue to respond in this way until some chance movement of a limb or part of the body happened to make contact with the lever thus opening the door. Repeated trials of the animal in this situation revealed that learning is a very gradual process taking many trials during which the animal begins to make fewer and fewer ineffective responses and the frequency of lever pressing gradually increases. The emphasis here is on the gradual and continuous nature of the learning which could be explained in terms of a strengthening of the appropriate response by reinforcement. Thorndike formulated the Law of Effect to account for such learning, namely,the responses that lead to favourable outcomes tend to be repeated.

Here then the criterion of learning was simply that the animal be successful in escaping from the box by whatever means and was capable of repeating such a blind trial-and -error solution when confronted with that same problem again. But other approaches were anxious to distinguish this trial-and-error learning from more sophisticated solution processes. For example psychologists in the Gestalt tradition, notably Kohler (1925) and Wertheimer (1945), used the term <u>insight</u> to describe a form of problem solving which yielded a problem solution that was not just a trial and error response, but rather was based on the subjects understandng of the structural characteristics of the problem.

An example of insightful problem solving often quoted is that of Kohler's chimp 'Sultan' solving the 'stick and fruit' task. In this problem the chimp is placed inside a cage outside of which is some item of fruit, say a banana. Naturally the chimp attempts to reach through the bars of the cage in order to retrieve the desired fruit but it has been placed too far away to be reached in this manner. However, inside the cage are two sticks. The sticks are of different diameter and hollow. The chimp readily uses one of the sticks as a tool with which to attempt to reach the banana but this is to no avail because each of the sticks alone is too short to reach the fruit. After many reaching attempts the chimp gives up the task. In Sultan´s case Kohler reports that later the same day Sultan was still in the cage idly playing with the two sticks when he chanced to put them end to end. Immediately Sultan "pushes the thinner one into the opening of the thicker, jumps up and is already on the way to the railings to which he has, up to now half turned his back and begins to draw the banana towards him".

In explaining these observations, Kohler argued that the critical factor was the animals grasp of the means-end relations that pertain among the elements of the task. The obtaining of the fruit was the goal of the problem, the attainment of which was blocked by the cage bars. This could be overcome by a combination of the two sticks not in a blind haphazard way but in a way decreed by the nature of the goal in relation to the obstacle.

While the naturalistic setting of these experiments and the rather anecdotal nature of the reporting is somewhat lacking in rigour, nevertheless these are interesting observations of what does appear to be quite spontaneous problem solving activity. Kohler uses these observations to point out the non-monotonic nature of this learning in contrast to the studies of Thorndike. Sultan, unlike Thorndike´s cats, seemed to arrive at a solution without any observable reference to presolution activity. Such learning cannot easily be captured by an associationist approach. Indeed as Yerkes points out:

> "If this particular type of solution should be exhibited under rigidly controlled experimental conditions by several specimens of chimpanzee few psychobiological observers would be likely to deny insight or to hesitate in describing the behavior as highly intelligent".

In the area of human problem solving Wertheimer distinguished between ´genuine´ and ´poor´ solutions on the basis of whether they involved an understanding of the problem´s inner structure. He viewed problems as consisting of an underlying organisation that decreed certain directions which a genuine solution must take. The perceptual analogy which he drew was the way in which a circle with a segment of its circumference missing tends perceptually to imply a ´closing off´ or a gap filling in the direction of completing the circle. The reason for this is that the circle was one of the preferred perceptual forms or ´good gestalts´ which have a privileged status in the perceptual system. To give a concrete example in the case of problem solving we might consider Wertheimer´s parallelogram problem. In fig 1.1 is a parallelogram of which we wish to find the area. For Wertheimer the parallelogram could be most 'productively' regarded as a deformed rectangle (cf. the dotted lines of figure 1.1), and furthermore, the deformation is such that the amount of area gained on the left hand-side is equivalent to the amount of area lost on the right-hand side. Consequently the area of the parallelogram must be equivalent to the area of the parent rectangle. The rectangle serves as a good gestalt for the parallelogram and the problem of finding its area is reduced to the simpler one of finding the area of a rectangle. Such an approach to the problem was regarded by Wertheimer as one which would lead to a genuine solution since it involved an insight into the inner structure of the problem.

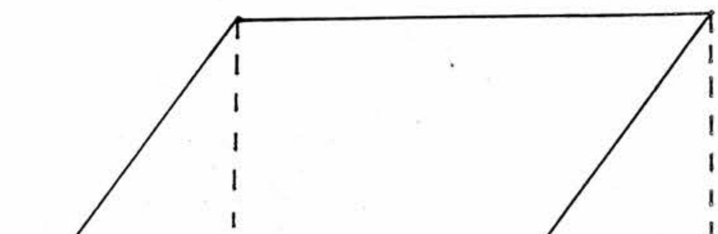


Fig 1.1

The Gestalt emphasis on 'genuine solutions' and the corresponding emphasis on the perception of structural relations, lead naturally to considering whether an 'expertise' factor might be involved in the distinction between trial and error and insight learning, and if so, what kind of expertise might this be?

But our two criteria of expertise namely that the solution be economical and transferable, are of little use here. Since Kohler made no empirical comparisons between learning by insight and trial-and-error learning in terms of say, solution time, it is difficult to compare the two forms in terms of their relative economy. Neither did he make any systematic study of the transferability of insightful solutions. It is not known for example whether Sultan went on to solve other stick and fruit problems more successfully or indeed whether he became more expert at ´insight problems´ in general. As for Wertheimer, while he maintained that transferability was the hallmark of ´genuine´ solutions he did not himself carry out any systematic studies of transfer. It is not known whether children taught the solution to the parallelogram problem went on to solve other geometry problems more successfully But while Kohler and Wertheimer did not themselves carry out such studies, there are transfer of learning studies which do bear directly on this issue.

Transfer Studies: Which kind of Expertise?

While there are many studies of transfer in the learning literature, only a small number of them have used a multiple problem transfer paradigm. But the work of Harlow et al (cf Harlow, 1959) is a particularly appropriate place to start since they attempted to explain ´learning-by-insight´ in terms of a gradual growth in expertise across problems.

Harlow (1949) used an object-discrimination task in which monkeys were presented with a pair of objects differing in multiple characteristics. The objects were presented on a tray and each of the objects covered a foodwell. The pair of objects were presented to the animal six times (in the original study) and it was required to choose one of the two objects. The position of the objects on the tray (i.e. over the left or the right food well) was varied in an unpredictable manner. Irrespective of its position, only one of the two objects was ever rewarded and it was the same object on each trial. When the animal had received the six trials with the first pair of objects then the second pair of objects, again differing in multiple characteristics and also different from the previous pair, were presented. The animal is again required to choose just the rewarded object.

Since the objects are unrelated to one another the only possible way for the animal to make the correct choice is to sample at random on the first trial and then, remembering which object it chose, move on to the next trial. If the animal was successful on the first trial then he must choose the same object on the second trial and if not he must choose the other object. Harlow presented his monkeys with up to 14 six-trial problems per day for several months and examined the relationship between intra-problem learning (ie. percentage correct on each of the six-trial pairs) and inter-problem learning (ie. percentage correct as a function of the number of previous pairs presented). Figures 1.2 and 1.3 summarise these results.

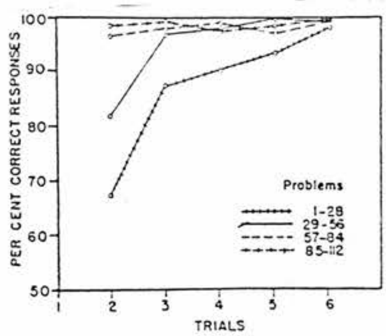


FIGURE 1.2 Discrimination reversal learning curves on successive blocks of problems.

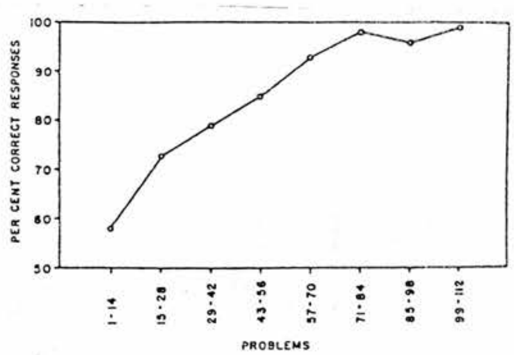


FIGURE 1.3 Discrimination reversal learning set curve based on Trial 2 responses.

The relationship between intra- and inter- problem learning is of interest here. As can be seen early on in the problems there is little inter-problem transfer manifest, learning is confined to improvement within problems. But with practice the animals' learning profile changes until eventually there is a high degree of transfer between problems, and effectively, one-trial learning within each of the problems. This can be seen clearly when performance on trial 2 of each problem is plotted against the number of problems solved, as in figure .3 above. Here after 80 problems or so, the animals solve a completely novel problem with almost 100% accuracy on only their second attempt at that problem, this is compared with little over 50% accuracy (chance responding) on their earliest problems.

Clearly these animals have developed a high degree of expertise in the solving of these problems, and this expertise has both intraand inter-problem components. As can be seen unlike the intra-problem component, the inter-problem component is essentially monotonic throughout. This was taken by Harlow to suggest that ´learning-by-insight´ could be explained in terms of transfer from previous problem experience.

This growth in expertise across problems was explained by Harlow in his **Error-factor Theory** (cf. Harlow 1952 for summary), according to which,the animal comes to a new task ready equipped with certain response tendencies. By systematic observation of the type of errors made by these monkeys Harlow et al were able to identify four such response tendencies. The first was Stimulus Perseveration wherein the animal was seen to exhibit a tendency to choose the same stimulus on successive trials irrespective of its reward value. The second was Differential Cue. Here the animal tended to respond to the the position of the rewarded foodwell rather than the changing position of the rewarded stimulus. The third type of error was termed Response Shift wherein the animal showed a tendency to respond to the non-rewarded stimulus after a series of successive choices of the rewarded stimulus. And finally there was the Position Habit tendency. Here the animal persistently chose the right or left object irrespective of the reward values. In this theory learning was viewed as the suppression of these error producing factors and thus the gradual emergence of the appropriate response profile which has been described as a Win-stay, lose-shift strategy.

As Mackintosh's (1974) short consideration of these findings points out, there are some difficulties with this theoretical explanation. It is not clear for example whether or not error factor elimination is the cause or the consequence of learning. As Mackintosh points out, on a six trial problem there are 2^5 i.e. 32 possible response sequences after the initial choice and error-factor elimination would account for only the elimination of a subset of these. What Mackintosh has failed to realize here though, is that Harlow's view of learning is essentially a hypothesis-testing view. Harlow is arguing that the animal is actively trying out various response sequences in a systematic and directed way and eliminating those which were ineffective. Consequently the animal would never consider the 32 possible permutations of response sequences. This would only occur if the animal was following a blind trial and error (with replacement) strategy.

What is really unclear about Harlow's error-factor theory however, is the status of the Win-stay, lose-shift' strategy that is the result of learning experience. Is this strategy simply part of the animals pre-existing repertoire of strategies that come into play when the simpler strategies such as response shift are proved ineffective? If so then clearly it is not the strategy that has been learnt in this situation. Rather, the animal learns that this task is of the type that calls for a certain class of response. Alternatively it may be that this new strategy is learnt in response to the failure of the animal's pre-existing repertoire, and presumably is then added to the repertoire. While this seems to be more congruent with Harlow's account of the learning set data it is quite clear that the error factor theory offers no account of the process by which such new strategies are forged from their predecessors.

Harlow´s work is a fine example of the use of a multiple-problem transfer paradigm as a means of generating some form of expertise in a problem solving task. Not only does the learning set experience allow Harlow´s monkeys to develop economical solutions to each problem, as is evidenced by the one-trial learning in later problems, but also it promotes the development of a solution which is transferable to novel problems. But how novel? What is not clear from this work is what kind of expertise is being developed. Is it domain-specific expertise, in which case transfer is limited to novel problems of the same class, or is it a more generalisable skill that has been developed?

Harlow certainly claimed to be studying a general problem solving skill, namely **learning-to-learn** (Harlow 1949), and he argued that the formation of learning sets;

> "Reduces the problem from an intellectual tribulation to an intellectual triviality and frees the organism to tackle problems of a new hierarchy of difficulty."

In this view it might be expected that the learning set animal was developing an expertise that not only allows it to solve two-choice discriminations effectively, but also allows it to deal with novel classes of problems more effectively. This being so, it might have been expected that Harlow would have gone on to study his animals on a new class of problems altogether, for example he might have transferred his animals onto an oddity discrimination task. Such a task might have served as Harlow's 'new hierarchy' of problems. But without such data it is difficult to know what kind of expertise is generated in these tasks. Some evidence from Shusterman (1962) however, indicates that the expertise may be much more domain specific than Harlow anticipated. In his study Shusterman found that monkeys trained on a **Win-shift, Lose-stay** strategy transferred badly to the more conventional learning set paradigm with its requirement for a **Win-Stay, Lose-Shift** strategy.

If indeed the learning set studies are demonstrations of domain specific expertise, it might be the case that more general transfer is simply beyond the bounds of the lower primates, and perhaps one of the unique characteristics of the human problem solver is a capacity for general expertise such as learning to learn, (McGonigle 1984). With this in mind we turn to a consideration of some of the transfer studies with human subjects.

Hull's work on concept formation offers a close analogy to the learning set studies of Harlow. This involved presenting subjects with a number of pseudo-Chinese characters (Hull 1945). These consisted of complex patterns of lines and curves. Of the set of characters a subset could be distinguished by the presence of one particular feature that was common to all characters in that subset. The characters presented to the subject were paired with nonsense syllables, all characters in the common subset being presented with the same nonsense word. Whether or not the subject had grasped the concept was assessed by whether or not he could name the members of the subset appropriately. The general results are well known, namely that subjects can indeed learn to associate a nonsense word with a particular stable element in a changing stimulus pattern and once established this can be extended to novel instances with that same element. Furthermore subjects can manifest this behaviour without necessarily being able to define what the common element is.

Hull´s original account of the learning process at work here, like error-factor theory, was essentially one in which irrelevant aspects of the stimulus became suppressed or inhibited. Here too as with extensions of the original learning set results, studies were made of the relative effectiveness of the number of trials per problem versus the number of problems solved. So for example, the studies of Adams (1954), on a task involving the learning of certain rules for the spatial arrangement of two objects, found that intensive training on a single problem produced more transfer to a novel problem than did multiple training. In contradiction to this, Hull himself found that moderate familiarity with half as many instances was more efficient than twice as thorough familiarity with half as many. This latter result seems more in keeping with the learning set studies (Harlow and Warren 1952, Callantine and Warren 1955) who found multiple training to be more effective.

This contradiction was taken up by Morrisett and Hovland in a single experimental study (1959). They argued that the crucial difference between the experiments of Adams and those of Callantine and Warren was the degree of mastery that subjects achieved within each problem. All of Adam's single-problem subjects mastered the training problem before going on to solve the test problems, the multiple-problem group did not, they were merely required to solve each problem once. From this they hypothesised that subjects who were given a multiple training paradigm but also allowed to master each problem before moving on to the next would show the most transfer.

This hypothesis was well supported by an experiment in which they compared three groups of subjects. One group were given 48 trials on a single problem before a test problem. The second group had 2 trials on each of 24 problems and group 3 had 16 trials on each of 3 training problems. Group 3 with a moderate degree of mastery on each of 3 problems showed the greatest transfer to the test problem despite the fact that it was hardly distinguishable from group 1 during the training phase. Group 2 showed an overall poor performance. While this study might have been even more convincing had a criterion of mastery been used, nevertheless the findings suggest, in accordance with Harlow's claims for the learning set, that expertise on a single problem is 'traded off' against expertise across multiple problems.

Morrisett and Hovland interpreted these findings in terms of an associationist theory of learning, but as Kendler (1961) points out, other studies, particularly those of Bruner, Goodnow and Austin (1956) attempted to identify some of the problem solving strategies subjects use in solving these types of problem.

The experiments of Bruner et al involved a concept formation task in which subjects were presented with cards picturing geometric patterns. The task for the subject was to find the principle that was guiding the experimenter's classification of some of these cards as members of a related set. As Bruner et al point out, there is a subtle distinction here between their approach and that of Hull, since for Bruner et al's subjects, the task was to identify the guiding concept rather than to use it to associate nonsense words with figures.

The subjects, after each presentation of a card, were asked to hypothesise a rule that they considered was plausibly in operation given the evidence so far. They were then informed if they were right or wrong and if wrong the experiment continued in the same way. The value of this approach is that rather like the learning set studies the nature of the subjects' errors is observed directly and consequently the decision-making process of the subject is to some extent externalised. Bruner et al found that there were regularities in this decision-making process, and these regularities they referred to as strategies.
To quote directly from Bruner et al;

"The phrase ´strategies of decision making´ is not meant in a metaphorical sense. A strategy refers to a pattern of decisions in the acquisition, retention and utilisation of information that serves to meet certain objectives, that is to ensure certain forms of outcome and to ensure against certain kinds of others. Among the objectives of a strategy are the following:

a) To ensure that the concept will be attained after the minimum number of encounters with relevant instances.

b) To be sure that the concept will be attained with certainty, regardless of the number of instances one must test en route to attainment.

c) To minimise the amount of strain on inference and memory capacity while at the same time ensuring that the concept will be attained.

d) To minimise the number of wrong categorisations prior to attaining a concept."

For Bruner et al, the strategy adopted by a subject is not fixed but depends upon the relative weighting given to the four objectives given above. And these in turn are determined by certain conditions. For example if the consequences of failure are severe then (d) becomes salient. If the nature of the instances encountered is such that retention is difficult then (c) becomes a priority objective. If there are time constraints imposed then (a) will be a priority objective and this will be to the detriment of (b).

Bruner et al identified two strategies at work. The first, termed the focussing strategy proceeds by first assuming that all the attributes of a positive instance define the concept and the subject responds accordingly. In the face of cards that are not instances of the concept, yet have some of the properties of the cards that are (ie. negative confirming cases) the subject does not change his hypothesis. Only when the subject is presented with a card that is an instance but does not share these properties does he change his hypothesis. In these cases the subject selects the subset of properties that are common to both instances and then proceeds. This strategy is mnemonically and inferentially economical, but requires large numbers of instances.

The second strategy that was adopted by subjects in Bruner et al's experiment is the scanning strategy. In this case the subject chooses only some properties of the initial exemplar. When this fails (ie positive and negative infirming cases) the subject seeks to change his hypothesis by reference to all the instances previously encountered. Consequently this strategy places a high demand on memory but can produce the correct concept in as little as one trial.

The study of Bruner et al is of interest because, like the learning set studies, it demonstrates how a multiple problem domain can be used to generate some form of expertise in the laboratory. Furthermore, Bruner et al offer some explanation as to the psychological mechanisms that underwrite such expertise, and the way in which the skilled problem solver might be able to assess the computational cost and benefits accruing to certain problem solving methods. Once again however, Bruner et al do not identify what kind of expertise is being developed here. Are the strategies identified by Bruner et al anything more than just local methods for solving a particular problem? The ability of subjects to assess the computational cost and benefit associated with particular problem solving methods would certainly seem to be a good candidate for a general problem solving skill. But like Harlow, Bruner et al did not study the performance of their subjects on other related tasks in which such skills might also be of use.

To conclude this review of the psychological literature then, it is possible to summarise as follows. While early studies were little concerned with expertise factors in problem solving, the research with both human and animal subjects using multiple problem training, leaves no doubt that expertise cannot only be observed in the laboratory but also that it is a significant variable in determining performance profiles. However, these studies have failed to identify what kind of expertise is being generated in these experiments, whether it is a domain limited form of expertise or a more general problem solving skill. The psychological research is simply silent on this issue. However there is research from the field of artificial intelligence which has made a quite clear distinction between these two forms of expertise. While there is little by way of experimental work in this field, and certainly there are no good studies of transfer, it does provide a framework within which an experimental programme can be evolved. This work is reviewed in the next chapter.

CHAPTER TWO

GENERAL MECHANISMS OR TASK SPECIFIC EXPERTISE?

Newell and Simon: The case for a General Problem Solver

Artificial intelligence views problem solving from a mathematical and formal perspective, consequently a problem solution is considered as something like a mathematical or logical proof and problem solving behaviour is seen as a search for such proofs in a given problem domain. As a consequence of this orientation the types of problems used in artificial intelligence are those that can be represented as a structured set of alternatives or options. The aim of the problem solver is to find an ordered sequence of such options that transform the initial state of the problem into the desired goal state, just as a mathematician might locate and put together the mathematical steps that constitute a proof of a theorem.

The now classic example of a general problem solving program is the ´general problem solver´ (GPS) of Ernst and Newell (1969). This program attempted to attain some level of generality by using a heuristic search technique which could be applied in a number of different problem domains in order to identify ordered sequences of options that would be appropriate for the generation of a solution in that problem domain. The rationale behind this attempt was that if many different problems can be represented as structured sets of options then it should be possible to develop quite general procedures for searching amongst options. The general search heuristic used by Ernst and Newell was called 'Means-End Analysis". Bundy (1978) cites the following informal example of the way in which a problem might be formulated in a way amenable to means-end search for a solution;

"My goal or end is to transform 'me at home' into 'me at Trafalgar Square'. The first task is to compare these two states to find the difference. I find the difference to be one of location. The means I have of reducing differences of location are operators like 'walk' or 'travel by train'. Some operators like 'walk' can be rejected as infeasible, but 'travel by train' is feasible, so my next task is to apply this operator to the initial state 'me at home'. Unfortunately the operator will not apply because the conditions are not right; I am not at the station. So I set up a new sub-goal to transform 'me at home' to 'me at station'. Again the difference is one of location and again I find the 'travel' operators. I can reject 'walk' as infeasible (I am lazy) and 'go by train' as a potential loop and select 'go by taxi'. This cannot be applied because the taxi driver does not know I need him. The difference is one of information, so I look for an operator that can reduce differences of information and find the communication operators like ´use the telephone!´...."

As can be seen from this example the essence of the problem solving process is the recursive use of operators to reduce differences between the current state and the goal state and this difference reduction can be carried out to any required depth. The result is a plan consisting of a sequence of operators each of which sets up the conditions necessary for the execution of the next operator in the hierarchy. Ernst and Newell were able to formulate a wide range of problems in such a way as to be amenable to means-end analysis. These ranged from mathematical puzzles such as the ´Tower of Hanoi´ and ´The bridges of Konigsberg´ through more complex mathematical problems involving integration to theorem proving predicate calculus and sentence parsing. Thus it would seem that GPS has indeed achieved some level of general problem solving capacity, but the work has not been considered as an unqualified success even from within the field of artificial intelligence research.

The cause for concern is that GPS requires an immense amount of task-specific information to be given to it before it is able to use the general search techniques effectively. As can perhaps be seen from the above example, for each new problem the initial state, the goal state and the relevant operators that can be used to transform these states have to be defined by the programmer. It is argued then that the actual contribution of GPS to the problem solution is quite small.

The problem highlighted here was recognised by Ernst and Newell at the outset. Their approach made a quite clear distinction between two aspects of problem solving, the formulation or representation of the problem and the heuristics or procedures that operate using that representation in a search for a solution. They argue that representational and procedural aspects interact in order to determine the generality of the system. To quote;

> "If we are to take seriously that generality is defined by the domain of problems that are solvable, then many perfectly general problem solvers exist: Turing Machines, algol compilers, etc.... but in describing the evaluation of a polynomial to a Turing Machine most of the problem solving techniques are contained in the specification of the problem....Thus the generality of a problem solver must be defined relative to the amount of information it takes to specify the problem.

It is clear then that in building a system of some generality the difficulties do not rest in designing procedures for solving a broad class of problems but rather in designing a representational system sufficiently powerful to allow a large class of problems to be defined with the minimum of task-specific information while at the same time being constrained enough to allow a finite set of procedures to operate on that representation.

Clearly, what is being suggested here is that there is a trade-off between the power of the system and the generality of the system, powerful problem solving procedures require a very sophisticated representational language, capable of representing a broad range of different problems in a similar format. Reducing the power of the problem solving procedures would put less strain on the representational device. But Ernst and Newell chose to hold the set of problem-solving procedures used by GPS constant, and in so doing it would seem that they must place most of the responsibility for generality into the representational domain, but it is here where the theory behind GPS has least to say;

> "..Nevertheless, the representation used in GPS was chosen ad hoc, within the framework of the problem solving techniques used, and not as a primary concern in implementing GPS."

It would seem however that unless there is an adequate theory of representation, this approach to generality is bound to fail. Without such a theory GPS is no better than a Turing machine with a Homunculus in the form of a programmer. It is because of this fundamental problem that most AI research on problem solving has now moved away from the GPS-type approach towards the expert systems paradigm in which programs are designed to produce efficient problem solving within circumscribed task-domains.

While the AI research has changed its emphasis from general skills to domain-specific expertise, there were some concerted attempts to use the GPS framework to study human problem solving, and here we might ask whether such a theory of human problem solving can explain how, if at all, the human system has overcome the engineering problem identified by Ernst and Newell.

The work of Newell and Simon (1972) was one such attempt, and the aim of the theory was to account for the problem solving behaviour in a number of problem domains by describing in computational terms the nature of the subjects´ representation of the problem environment and the procedures used in attempts to find solutions. They argued that a three-way distinction must be maintained in any analysis of problem solving, between the structure of the task environment, the subjects representation in that in environment, which may differ from individual to individual, and the problem solving procedures that are used to solve the problem.

As a consequence of these distinctions it was necessary for Newell and Simon to have some means of accessing the nature of the problem representations or ´problem space´ that a given subject adopts in his attempt to solve a problem. To this end they used the technique of verbal protocol reporting, which involved the subject in providing a 'running commentary' of his thoughts as he proceeded to solve the problem. Such 'think aloud' protocols were not considered as introspection but as behaviour that could legitimately be seen as tracing the subject's path through his representation of the task environment or 'problem space'. To quote;

> "This is not a space that can yet be pointed to and described as an objective fact for the human subject. An attempt at describing it amounts, again, to constructing a representation of the task environment - the subjects representation in this case. The subject is presented in the experiment with a set of instructions and a sequence of stimuli. He must encode these components - defining goals, rules and other aspects of the situation - in some kind of space that represents the initial situation presented to him, the desired goal situation, various intermediate states imagined or experienced, as well as any concepts he uses to describe these situations to himself."

It may be noticed that the latter part of this quote which describes the contents of this problem space sounds much like the type of description used above as an example of means end analysis in GPS with its emphasis on goals, initial states and intermediate states. And indeed it is the case that Newell and Simon used information obtained from protocol data to compare subjects´ search through their problem spaces with the trace of GPS solving the same problems. This was achieved by an intermediate level of description called ´the problem behaviour graph´. This interpreted the subjects´ utterances in terms of states and operators and described the choices between operators that subjects´ were making. This was then compared with the choices that GPS made in the same situations.

There is no doubt that this method provides evidence for subjects´ use of states and operators and their use of means-end analysis which is entirely consistent with the GPS model. As Bundy points out there are some differences of detail, for example human subjects tend to use more elaborate back-up in response to error and sometimes use more sophisticated means of eliminating infeasible operators, but by and large the fit between model and human protocol data is good. As such GPS clearly has some claim to being a ´psychologically plausible´ model of problem solving. But this fit between the GPS and the human data in the light of the criticisms levelled at GPS, only serves to raise questions concerning the generality of the human problem solver. Is there any evidence in Newell and Simon´s account of human problem solving of a level of generality beyond that attained by GPS, and if so how does the human problem solver overcome the computational problems identified by the GPS research?

In order to address this issue, what would be required is an understanding of the relationship between the task environment and the problem space. That is to say an account of the factors that influence subjects´ representation of the problem solving task. While Newell and Simon make it perfectly clear that the relationship is not a simple one, and that for any given task environment there are a large if not infinite number of possible problem spaces, they do not offer any theoretical account of how a problem space is chosen and constructed. The use of verbal protocols as the sole behavioural data means that their analysis of problem solving only begins once the problem space is constructed. They do not experimentally manipulate variables that might have conceivably affected such a construction process. Thus as with the GPS model upon which it was based, Newell and Simon´s theory of human problem solving is at its weakest at the point of most stress. By Newell and Simon´s own admission;

> "The theory of problem solving to be presented in this book has much more to say about methods and executive organisation than about creating new representations or shifting from one representation to another. "

Thus it would seem that support from the field of artificial intelligence for a belief in problem solving skills of some generality is not forthcoming. The questions of generality raised at the beginning of this chapter in the light of the learning set research receive no support from this new approach. In his chapter 'Problem Solving and Education', Simon offers the following conclusion concerning general problem solving skills;

> "There is some evidence that problem solving skills can be taught, although there is regrettably little evidence that such instruction is cost effective as compared with equal effort devoted to subject matter skills."
>
> (Simon 1980)

This would seem to suggest that Simon himself has little faith in the notion of task-independent problem solving skills, and that further attempts to examine such skills, if they exist, might be fruitless. Thus we shall set aside such considerations for tha moment and turn instead to the alternative characterisation of expertise, namely, as a domain-specific skill, and examine how this has been viewed from the area of AI and related work.

Domain Specific Expertise and State Space Analysis

A study of domain specific expertise must begin with an analysis of the task domain. Such an analysis of problem classes must be independent of psychological variables such as transfer, because if not there is a risk of providing only a tautologous definition of the task domain in terms of the set of problems over which transfer is manifest, while at the same time claiming the transfer to be domain specific. But at the same time if this formal analysis is to be of any use it must be assumed to have psychological reality at some level. It is in providing such an analytical framework that AI research is so useful.

One formalism that has been used to this effect in AI research is termed state space analysis (Nilsson 1972, Luger 1978, Goldin and McClintock 1980). For a large class of problems, (certainly for all those used by Ernst and Newell for eg.) it is possible to define the structure of those problems in terms of states and relations between states. Board games such as chess and draughts for example can be described as a finite set of legal configurations of the pieces on the board, and in a state-space analysis this would correspond to the set of states in a state space graph. Furthermore the legal moves in the game which transform one configuration into the next are represented as arcs in the state space graph. Given the exhaustive set of legal moves and the resulting states of play it is possible to characterise the play of the game as a path through the state space.

This type of analysis can be used to define the relationship between two different problems and to define classes of identical problems. Two problems are said to be identical or isomorphic when they are described by the same state space graph. To give a concrete example we turn to a consideration of the problem that has been used extensively in AI research on problem solving; the Tower of Hanoi (ToH) problem.

This problem comprises of a board upon which stand three pegs. At the start of the problem there are four disks of regularly decreasing sizes stacked up on the lefthand peg with the largest at the bottom. This is illustrated in figure 2.1 below.

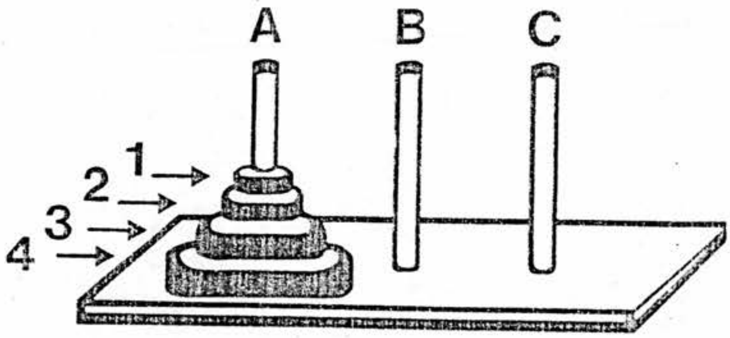


Fig 2.1.

The Tower of Hanoi problem in its start position.

The task for the problem solver is to transfer the disks, moving only one at a time from the lefthand peg to the righthand peg. But in doing so he must never allow a larger disk to stand on top of a smaller disk. This constraint defines the legal configurations of the disks on the pegs and thus the legal moves that the subject can make, and thus it is possible to describe the state space graph for this problem. This graph represents the problem as a branching tree of nodes and arcs in which the nodes are the legal configurations of the disks on the pegs and the arcs are the permissible transformations of the configurations. Figure 2.2 presents this graph;

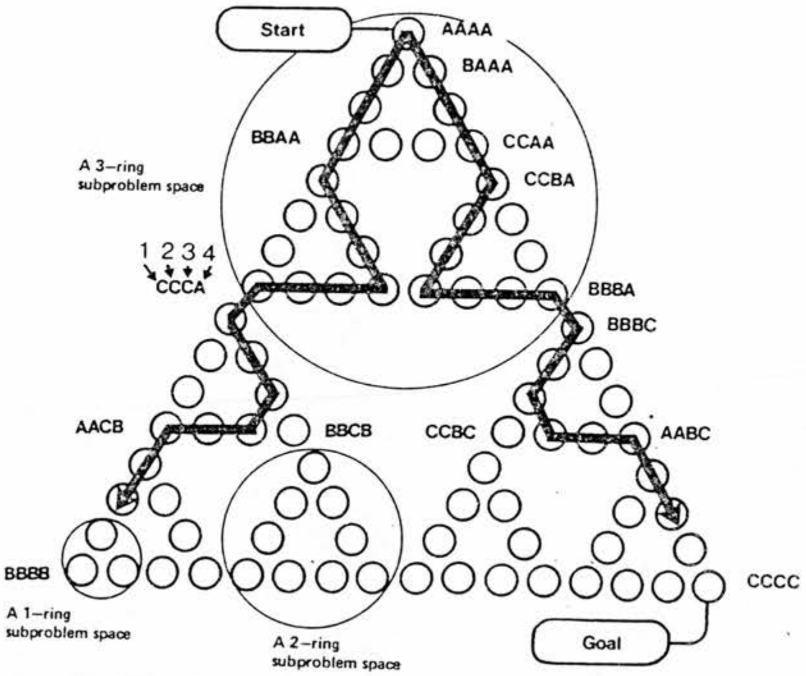


figure 2.2 The state space graph for the ToH problem showing 1-, 2- and 3- ring sub problem spaces.

The configurations are seen as labelled nodes with the initial state or start state of the problem as the topmost node. The label of this node is AAAA. These 4 letters indicate that the four disks are all on peg A - the left hand peg- at the start of the problem. The leftmost letter A denotes the position of the smallest disk, the rightmost A the position of the largest disk, and so on.

From the initial state of the problem only two legal moves are possible, the subject can move the smallest disk to peg B, or to peg C. These options are represented by the two nodes immediately below the topmost node of the graph. The label on the righthand side; BAAA, indicates that the smallest disk can be moved to peg B, the lefthand node. CAAA, indicates the only other option of moving the smallest disk to peg C. In this way all the move options from each successive state of play can be described. The bottommost righthand node is labelled CCCC, indicating that all the disks are on the righthand peg. This, it will be recalled, is the goal state, the attainment of which means the problem has been solved. Consequently, it is possible to see from the state space graph that the shortest solution path is to carry out that sequence of 15 moves represented by the righthand side of the graph. This constitutes the minimum solution path (MSP) for the ToH problems.

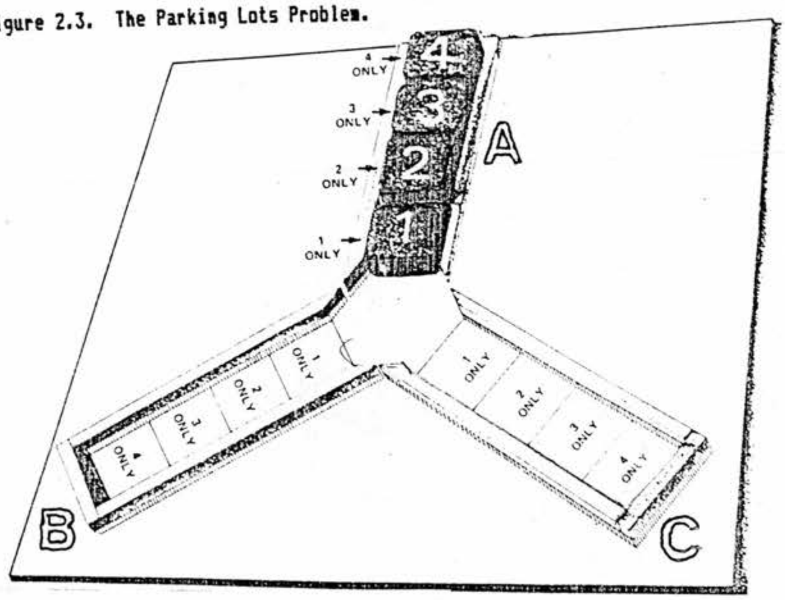
There are a number of other structural characteristics of the ToH problems that become apparent from the state space graph. Most notable is the highly recursive substructure and symmetry of the problem. The ToH problem is a recursive problem in that in order to solve the 4-ring problem, the subject must first solve a 3-ring problem, and in order to solve a 3-ring problem the subject must first solve a 2-ring problem and so on. That is to say, in order to solve the 4- ring problem the subject must first get the biggest disk to the goal peg. In order to do this it is necessary to have all the other three disks stacked appropriately on the intermediate peg - peg B. The subject must thus solve the 3-ring problem for peg B before solving the 4-ring problem for peg C. In order to get the smallest three disks to peg B the subject must first get the largest of the three to peg B. In order to do that he must first solve the 2-ring problem for peg C, and so on.

This recursive substructure is readily seen in the state space graph as a nested set of pyramids of nodes and arcs. The pyramidal shape of the graph as a whole represents the 4-ring problem, but that pyramid is itself made up of 3 smaller pyramids each of which is a self-contained 3-ring version of the ToH problem. These in turn are each made up of the pyramids representing 2-ring problems. A 3-ring and a 2-ring subproblem are circled in figure 2.2. This substructure produces a symmetry in the ToH problem which again is obvious from the graph. The moves represented by the lefthand side of the pyramid are a reflection of the minimum solution path of the righthand side but where the goal peg, instead of being peg C is peg B and so on. It is this highly recursive structure that makes this problem so amenable to the kinds of recursive means-end techniques embodied in AI models like GPS since the same move patterns are present at each level of the recursion.

Of course, the state-space analysis is only one of possibly many formalisms for the analysis of these problems, Nilsson (op cit) for example provides at least one other approach. But why this formalism is of particular interest is that there is reason to believe that this level of analysis has some psychological reality.

Luger (1980) argues that it is the detection of these structural invariances and symmetries that constitutes the major problem for a human problem solver. It is by the systematic exploration of these move structures he argues, that a subject comes to solve the ToH problem and that this exploration can be quantified by plotting a subjects path through the state space graph. If this is done some systematic characteristics can be recognised. For example, the MSP from peg A to peg C will often be preceded by a symmetrical MSP but from peg A to peg B. Likewise once a subject has solved a 3-ring subproblem in a MSP other 3-ring subproblems can be solved in a MSP also. Now consider the following problem;

Figure 2.3. The Parking Lots Problem.



Numbers in diagram correspond to colours in text

The problem is described as follows;

"On the board in front of you are four delivery trucks from four rival companies. They have been travelling to town along road A. They must go to the factory in road C, which, to simplify loading and unloading, has been divided into parking spaces. Each company has its own company-owned parking lot and the colours of the lots correspond to the company colours on the lorries themselves. There are similar company-owned parking lots in roads A and C. As can be seen, the roads are narrow and no overtaking is possible, and no lorry may park in the central cross-hatched area as this causes congestion. A final problem is that each company zealously guards its own company lots and will not allow lorries from rival companies to park on them even momentarily. Given these difficulties can you show how to manoeuvre all of the lorries into their appropriate lots in road C so that they might be able to deliver their goods to the factory."

It is possible to plot a state space graph for this problem also. From the initial state one must either move the first lorry to its parking lot in road C or road B. From there it would be possible to move the second lorry to the other road and so on. If the state-space graph of the parking lot problem is constructed in this way then despite the apparent differences between it and the ToH problem discussed above, the two graphs turn out to be identical. This identity between the two problems can be seen if the 4 lorries are replaced by the 4 disks of the ToH problem and the 3 roads by the 3 pegs, and the parking constraints by the stacking constraints on the disks of the ToH. Because of this structural identity between the two problems they are said to be <u>isomorphic</u> modulo their state space decomposition.

Thus we can see how the state space analysis of problems allows us to define classes of structurally equivalent problems. Formally speaking, while such problems vary immensely in their surface structure characteristics, a solution to any one of these problems that is based on subproblem invariances can stand as a solution to any problem from that structurally defined class.

With this formal taxonomy of problems then, it is possible to make two claims about domain specific expertise. Firstly, it would be expected that an expert solution will be one that exploits the structural invariances inherent in these problems. Second, such expertise should guarantee that a problem solver will be able to transfer his solution to other structurally isomorphic problems. There are in fact three experimental studies of transfer using state space tasks that have attempted to assess these claims. These are reviewed below with respect to what evidence they offer for the existence of domain specific expertise.

Some Transfer Studies with State Space Tasks

The first study to be reviewed is by Luger and Bauer (1976). This study used two state-space problems, the ToH described above and an isomorphic variant of it called the tea ceremony (TC). This task although differing somewhat in its description employs a board essentially the same as the parking lot problem described above. Appendix A includes a full description of this task along with all the other state space problems referred to in the thesis.

In this study two groups of students solved both of these tasks in a ´train and test´ procedure. One group solved the ToH problem first, followed by the TC problem while the other group solved the problems in the reverse order; ie. TC followed by ToH. The hypothesis under test was that if solution to these problems was underwritten by an exploitation of the subproblem invariances described by the state space graph, then since both problems have the same invariances subjects should transfer positively across the two problems.

Luger and Bauer measured total time and total number of moves to solution for the two problems. An analysis of variance of these data revealed the following results. There was no significant difference between the groups suggesting that the order in which the problems were solved does not affect overall solution time or moves -to- solution. There was no overall effect of practice suggesting that the difficulty of the problem solved second by each group was not reduced by prior experience of the other isomorphic problem.

While this result appears to refute the hypothesis of Luger and Bauer there was a significant effect of problem type, which suggests that one of the problems, the TC problem, requires significantly longer time and greater total moves for its solution. Fortunately for the hypothesis this effect of problem type interacts with problem order for both moves and time, suggesting that this difference between the two problems is reduced as a function of practice. This implies that some transfer between the two problems has indeed occurred. While Luger and Bauer did not carry out any post-test for the location of this interaction, by inspection of their summary tables it would seem that solving the TC problem significantly reduces the difficulty of solving the ToH subsequently.

Thus it would seem from this study that there is some evidence for transfer across isomorphic problems. However, this experiment, as it stands, tells us very little about the nature of the expertise that underwrites such transfer. This is so for two reasons. Firstly as was made clear in the review of psychological studies a train and test paradigm of the type used here is not particularly useful for an examination of expertise. It is not known whether subjects in solving their first problem did in fact use the structural invariances of that problem to produce an expert solution as distinct from solving the problem by some form of trial and error learning.

But even if some level of expertise had been generated in problem 1, the second objection would still hold, namely with the experiment as it stands, there is no way of knowing whether this expertise was genuine domain specific expertise of the type envisaged by Luger, or whether it was simply some form of general problem solving skill such as means-end analysis of the type used by GPS to solve just this type of problem. Because Luger did not include a control condition testing for transfer across <u>non-isomorphic</u> problems he has not differentiated between the two forms of expertise.

The second study (Reed et al 1974) was essentially identical in design to that of Luger but using a different kind of state-space problem. The first of these problems was the 'Missionaries and Cannibals' task, in which the subject is required to transport three missionaries and three cannibals across a river using a boat which holds two of them at most. The problem arises because if at any time the cannibals on either riverbank outnumber the missionaries on that bank, then the cannibals will indulge in their habit of eating missionaries.

The state space of this problem (see appendix A) is much simpler than that of the ToH, it consists of only 14 states and the solution path is 11 moves long, consequently the choice points for possible legal moves is very limited in comparison to the ToH problem. While there is some symmetry about the centre state indicating that the first half of the solution path (getting all 3 missionaries onto the far bank) is the mirror image of the second half (getting all 6 people onto the far bank), there is very little of the rich subproblem structure found in the ToH problem.

The second task used in this study is called 'the Jealous Husbands' problem. In this task 3 husbands and their wives replace the missionaries and cannibals. The problem arises because the 3 husbands will not allow their wives to be alone with any of the other men. When the conventional state space graphs of these two problems are compared they can be seen to have identical structures. If however the state spaces are extended to include not only legal but also illegal moves then the state spaces are distinct, the jealous husbands (JH) problem having more illegal configurations than the missionaries and cannibals problem (MC). This difference arises because in the JH task there is an additional pairing constraint on the wives and husbands, for example a wife would not be allowed in a boat with a man who is not her husband, whereas in the MC task any cannibal may share the boat with any missionary. Because of this distinction the two problems are said to be homomorphic rather than isomorphic.

Reed et al used a ´train and test´ procedure like that of Luger and Bauer; one group solved the MC problems followed by the JH problem and the other group solved the reverse order. They measured moves -to- solution and solution time as well as the number of illegal moves the subject made. Unlike the Luger and Bauer study however they found no evidence of transfer effects for moves or time to solution, there was however some very minor reduction in the number of illegal moves made, but they do not accept this as strong evidence of transfer.

In a follow-up study Reed et al looked at transfer within the same problem (either JH or MC) solved twice. They found that for the JH problem there was a significant reduction in time and illegal moves to solution between the first and the second attempts at the problem but no effect for the number of moves to solution. For the MC problem there was a significant reduction between the first and second attempts only for solution time. Reed et al take this as evidence that the failure to find transfer between problems in their first experiment was attributable to the failure of subjects to recognise problem relatedness.

So in their final study the 'train and test' paradigm of experiment 1 was again used but this time subjects were given explicit information about the relatedness of the two problems before they moved on to solve the second problem. The results revealed a significant transfer between the two problems but this was confined to the group solving the JH problem first followed by the MC problem. This effect was significant for time and illegal moves only.

The failure to obtain any clear evidence for transfer here seriously weakens Luger's claims about the psychological reality of state space invariance. Once again however it is not clear whether failure to find strong transfer is due to the use of the train and test paradigm. It is not known for example, whether the subjects had mastered the first problem before going on to solve the second. What might have been particularly useful here, is the use of some form of criterion of mastery paradigm in which subjects are required to solve each of a large set of probems to some level of mastery before going on to attempt further problems. In this way it would have been possible to ensure expertise within a single problem before testing the generality of such expertise. Ironically, in the third experiment to be reviewed here, Thomas (1974), uses an isomorph of the MC task to examine <u>intra-problem</u> transfer, but he does not go on to examine <u>inter-problem effects.</u>

In the problem used by Thomas, hobbits replace misionaries and orcs replace cannibals. He used two groups of subjects, one group - the control group, solved the problem from the beginning to the end once only. The experimental group however began by solving the last half of the problem from the middle-state to the end state, and only then went on to solve the whole problem from beginning to end. Thomas argued that if moves through the state-space was the appropriate level of analysis, then subjects in the experimental group solving the second half of the problem twice should show an improved performance on their second attempt.

However, the mean total moves required to solve the halves of the problem were compared to reveal that the experimental group not only solved the <u>first</u> half of the problem in fewer moves then the control group but they also solved the second half in fewer moves than the controls, even on their <u>first</u> attempt. Furthermore the experimental group showed no improvement on their second attempt at the second half of the problem.

Thus it would seem that while there is some evidence of transfer for the experimental group between halves 1 and 2, there is no effect on a second attempt at the second half. Furthermore there appears to be negative transfer from the first half to the second half for the experimental group. As was mentioned above, there is some symmetry between the first and second half of the problem which might explain the transfer between these two halves for the experimental group, the negative transfer of the control group and the failure to obtain transfer where it was expected across the first and second attempts at the second half remains a mystery.

Thomas concludes that there is no evidence for the psychological reality of state space analysis, but as with the other two experiments reviewed here, weaknesses in the methodology temper such conclusions. The failure to ensure that subjects were not merely <u>successful</u> on their first attempt at the problem but had actually mastered the problem more critically means that there are only the slimmest grounds for claiming this experiment as evidence against the role of state-space structure in determining transfer effects. It is clear that these experiments do not allow for any firm conclusions to be drawn about expertise factors in the solving of these problems.

In conclusion then, work from the field of AI has provided us with analyses of both general problem solving skill and expertise of a more domain specific sort. On formal grounds alone, there seems to be a strong case for believing that general problem solving expertise may be very much more difficult to analyse in psychological terms than expertise of a more domain-specific sort. On the other hand the attempts that have been made in this area to investigate domain specific expertise experimentally, have failed to meet even minimal methodological standards set by the best psychological research on transfer.

This does not mean to say however that the AI approach must be abandoned. On the contrary, this research has provided us with a very usable framework for analysing problem structure formally, and thus defining problem classes, without which a study of domain specific expertise cannot get off the ground. Rather than give up this framework, it would seem more prudent to import some of the wisdom of the psychological studies of transfer into this task domain, and to use a synthesis of psychological and formal approaches as a lever to tease apart the two notions of expertise which have been so persistently confounded in the experimental literature. The experiments reported in the chapter that follows are just such an attempt.

CHAPTER THREE

AN EMPIRICAL EVALUATION OF STATE SPACE TRANSFER

This chapter reports two experiments with state space problems designed to assess whether, under proper experimental conditions, it is possible to generate some form of expertise in this problem domain, and secondly to investigate what kind of expertise this might be.

Experiment 1: Isomorphic and Non-Isomorphic Transfer Compared.

In the first experiment subjects are presented with 8 state space problems to solve. On the basis of a state space analysis, these 8 problems form two sets. 4 of the problems are isomorphs of the ToH and hence all have the same state space structure. Thus each of the 4 problems in this set are of the same type modulo their state-space analysis. The other 4 problems are not of the same type. Each of the problems in this set has a different state-space structure. In examing transfer effects in this multiple problem domain it is hoped that two things will emerge.

Firstly, by using this somewhat larger number of problems than has previously been used, clear evidence of some form of transfer effects will emerge. These will be indexed by a reduction in solution time and moves to solution. Secondly, in examining the domain of problems over which such transfer is observed, it is hoped that it will be possible to infer whether such transfer is underwritten by domain specific expertise as Luger envisages, or whether some more general form of expertise is involved. If it is the former, we would expect transfer effects to be limited to the isomorphic set of problems. If the latter is involved, we would expect transfer effects to extend over both sets of problems.

Method

The Problems

It was felt that 4 problems per set would allow a sufficiently detailed transfer profile to emerge, consequently 4 isomorphic versions of the Tower of Hanoi (ToH) were designed. These will hereafter be referred to as the Iso-set. A contrastive set of Non-isomorphic problems were also designed. This set, hereafter referred to as the Non-iso set comprised of the Missionaries and Cannibals problem (MC), a version of the two pails problem called the Sugar Refineries (SR), a version of the 3-Nickels and Dimes problem called the Shuntyard (SY) and a version of the farmyard problem (FY). The original versions of all of these problems are described in Goldin and McClintock (1980) and details of the particular tasks used here are included as Appendix A. All of the non-iso problems are distinct modulo their state-space decomposition and all are distinct from the ToH isomorphs. All of the problems are one-person problems and were presented on wooden boards along with the necessary story context.

Subjects

The subjects for this experiment were 1st-year undergraduates attending Psychology I lectures at the University of Edinburgh and who volunteered to act as subjects. They were not paid for their services. Subjects of both sexes took part in the experiment. There were 16 subjects, 8 in each of the experimental groups.

Design and Procedure

The two 4-problem sets were presented to both groups of subjects but the order of presentation of the two sets was counterbalanced across the two groups as in table 3.1.1a. The order of presentation of the 4 problems within each of the problem sets was also counterbalanced for each of the groups. Thus for each cell of table 3.1.1a the order of presentation of the problems was counterbalanced as in table 3.1.1b. Consequently of the 4 problems in each set each problem was solved in each of the 8 possible positions of presentation by some subjects.

Table 3.1.1a. Design of Experiment 1

	first problem set	second problem set
Group 1	ISOMORPHIC	NON-ISOMORPHIC
Group 2	NON-ISOMORPHIC	ISOMORPHIC

Table 3.1.1b. Counterbalancing of problems within cells.

subgroups	order of presentation			
n=2	P1	P2	P3	P4
n=2	P2	P3	P4	P1
n=2	P3	P4	P1	P2
n=2	P4	P1	P2	P3

Subjects were required to solve each of the 8 problems once only and were encouraged to solve the problems as quickly as was consistent with not making unnecessary errors. Subjects were presented with the problems one at a time and were not told how many problems they would be asked to solve. Before commencing each problem the subject had the story context associated with that problem read to him and then read the story for himself. The subjects were not informed about the formal characteristics of the problems. During each subject's problem solving the experimenter recorded the number of moves made by the subject as well as the time taken to solve each of the problems.

Results

The mean solution times for both groups solving each set of problems (ie. the means corresponding to the cells of table 3.1.1a) are presented in table 3.1.2.

	first problem set	second problem set
Group 1(I,NI)	162	182
Group 2(NI,I)	326	150

Table 3.1.2 Comparison between groups of mean solution time (seconds) for each problem set.

An analysis of variance on these data (BMDP P2V) revealed a significant effect of group ($F_{1,14} = 6.64$; $p = 0.0219$) and problem set ($F_{1,14} = 6.19$; $p = 0.0261$), and a significant group by set interaction ($F_{1.14} = 9.61$; $p = 0.0078$), and a further analysis of group effects for each set revealed a significant effect of group for the first set only ($F_{1,14} = 10.80$; $p = 0.0054$), there was no difference between the groups on the second set of problems ($F_{1,14} = 1.26$; $p = 0.2803$). These results suggest firstly, that the non-iso set considered as a whole when solved by naive subjects is significantly more difficult than the iso set. However it appears that experience of solving the iso set significantly reduced the difficulty of the non-iso set when solved subsequently. This is supported by further analyses of variance which showed a significant difference between problem sets when solved by group 2 (non-iso -> iso), ($F_{1,7}$ = 13.93; p = 0.0100), but not when solved by group 1 (iso -> non-iso) ($F_{1,7}$ = 0.27; p = 0.6101). Thus there is an asymmetry between the two groups. Experience on the iso set transfers positively to the solution of the non-iso problems but the converse is not true, experience on the non-iso problems does not appear to transfer to the iso set.

Turning now to examination of transfer effects within each of the cells of table 3.1.2a, figure 3.1.1 presents the mean solution times to problems within each of the the cells of table 3.1.1.

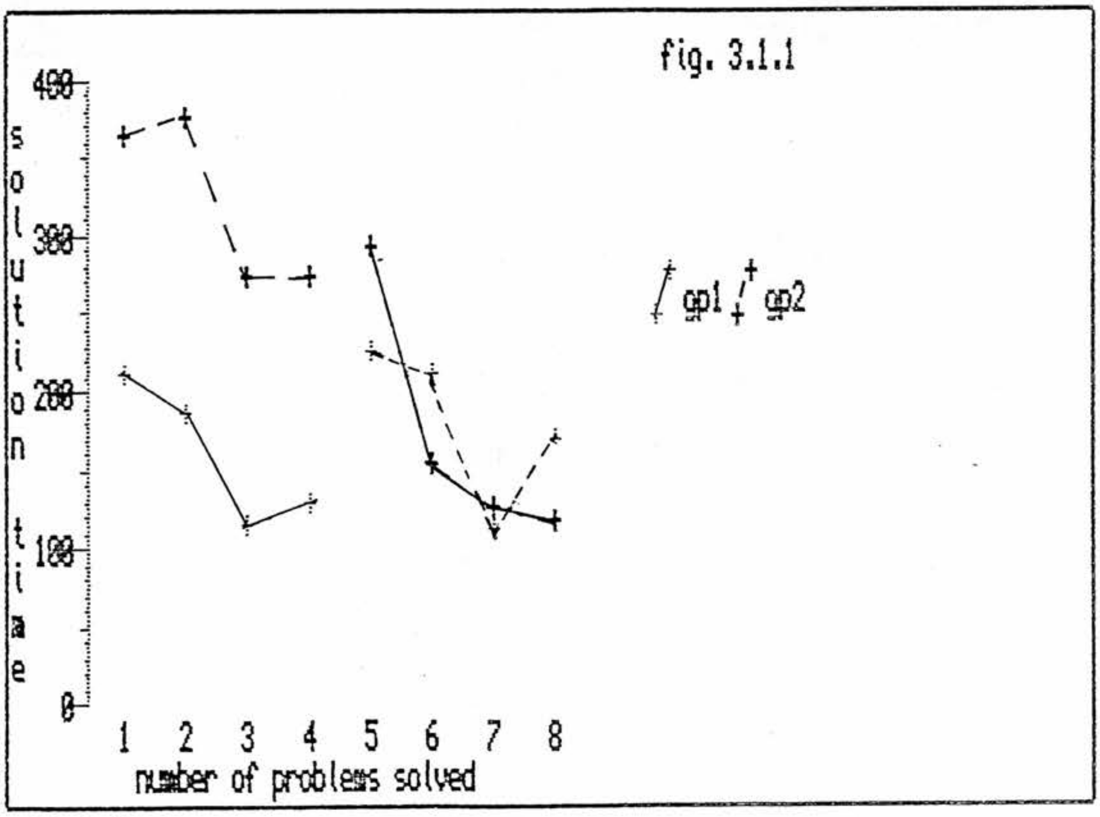


Figure 3.1.1 Solution Time (secs) as a function of number of problems solved. Problem 5 marks the start of a new problem set for both groups (non-iso's; gp.1, iso's; gp 2)

While it appears from an examination of this figure that there are quite large inter-set effects, in fact an analysis of variance on these data reveal no significant effect of practice within each set ($F_{3,21} = 1.77; p = 0.1839$). This suggests there is a large amount of variance attributable to the individual problems that go to make up the means for each set. Figure 3.1.2 a&b, presents an analysis of mean solution times as a function of position of presentation for each of the 8 problems that make up the two problem sets.

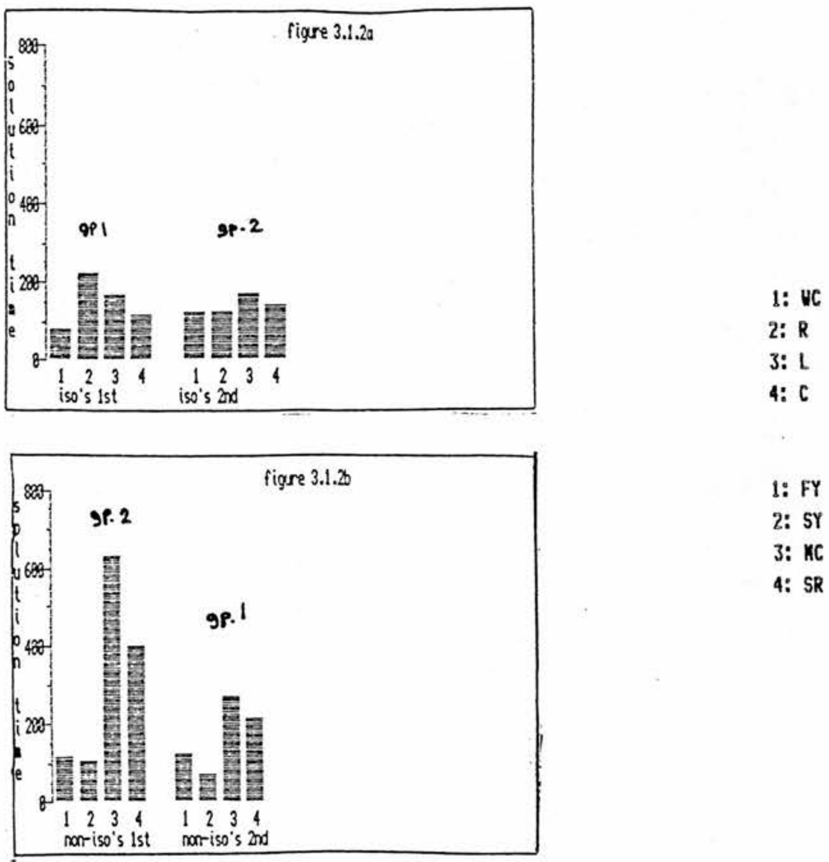


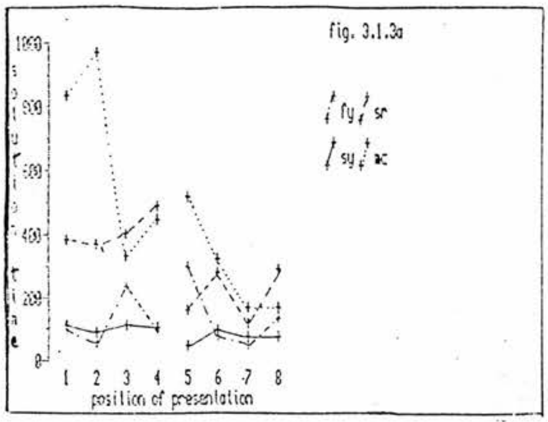
Figure 3.1.2 Solution times (secs) for the individual problems making up the ISOMORPHIC set (3.1.2a) and the NON-ISOMORPHIC set (3.1.2b), as a function of position of presentation.

As can be seen there are striking differences between the problems and an analysis of variance revealed a significant effect of problem ($F_{3,18}$ = 13.8; p < 0.0001), a significant effect of set position ($F_{1,6}$ = 79.3; p < 0.0001) and a problem by set position interaction ($F_{3,18}$ = 10.3; p = 0.0004) suggesting that the large differences between the individual problems is reduced as a function of practice. A Scheffe test for post-hoc comparisons (Hayes 1976), indicates that the effect of problem is attributable to one problem alone - the MC problem (problem 3 in fig. 3.1.2). This problem is more difficult than all other problems at set position 1 (p < 0.05), but this difference is lost at set position 2. In fact, at set position 2 no single problem is significantly more difficult than any other.

These results suggest that the inter-set effects of group 1 can be re-interpreted as a progressive decline in the difficulty of one problem alone (the MC problem) as a function of practice. This interpretation is supported by a consideration of the solution times for each of the 8 problems as a function of position of presentation as in figures 3.1.3 a&b (overleaf)

As can be seen in contrast to the initial suggestions of no inter-set effects it can be seen that there are quite dramatic transfer effects for certain problems. In particular the MC problem shows a quite marked monotonic reduction in solution time as a function of position of presentation. T-test on these data for the MC problem alone show a significant difference in solution time between positions (1+2) versus (3+4), (p < 0.05) but not for

Figure 3.1.3a. Solution time (seconds) for the individual problems that make up non-isomorphic problem set as a function of position of presentation of each problem.



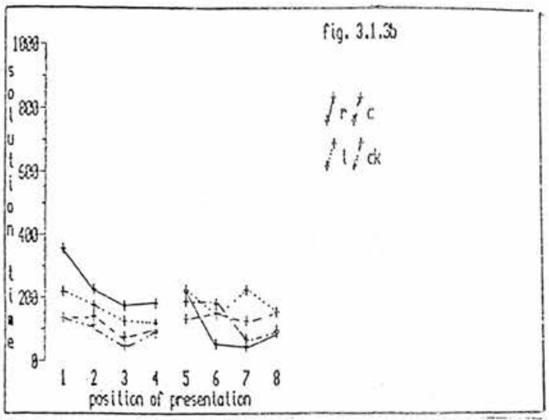


Figure 3.1.3b. Solution time (seconds) for the individual problems that make up isomorphic problem set as a function of position of presentation of each problem.

positions (5+6) versus (7+8), ($p > 0.05$). A simple linear regression shows this decrease in solution time for the MC problem to be linear across positions ($r = 0.73$; $p < 0.05$). There are no similar effects for the only problem approaching MC in difficulty namely the SR problem, (although fig. 3.1.3a does show some indication of inter-set effects for this problem). This suggests that the transfer obtained in this experiment is limited to a single difficult problem and that the best predictor of transfer is simply the number of previous problems solved irrespective of the structural characteristics of those problems. As can be seen from figure 3.1.3b there are no similar effects for the iso problems. While experience on the iso set seems to improve performance on the non-iso MC problem there appears to be no improvement across the iso- set as would have been predicted by Luger. The one possible exception is the relatively difficult 'rockets' problem which showed some (non-significant) reduction over the iso set. It would appear that the Formal taxonomy of Iso versus non-iso is not useful for predicting either task difficulty or transfer effects as indexed by solution time. A consideration of moves- to -solution provides a similar profile.

Since the minimum number of moves to solution varies amongst the eight problems used in this experiment, a corrected moves total was used in the analysis. This was derived as the difference between the actual moves-to-solution observed and the length of the minimum solution path for a given problem. Thus a subject who solved the ToH problem in 25 moves would obtain a corrected score of 25-15, ie. 10.

Figure 3.1.4 (overleaf) presents the corrected moves total as a function of problem set and number of problems solved within each set. As can be seen there is little evidence for group differences or transfer and an analysis of variance revealed no significant effect of group ($F_{1,14} = 0.04$; $p = 0.8438$) or problem set ($F_{1,14} = 1.81$; $p = 0.2004$) or practice within each set ($F_{7,98} = 0.88$; $p = 0.5221$).

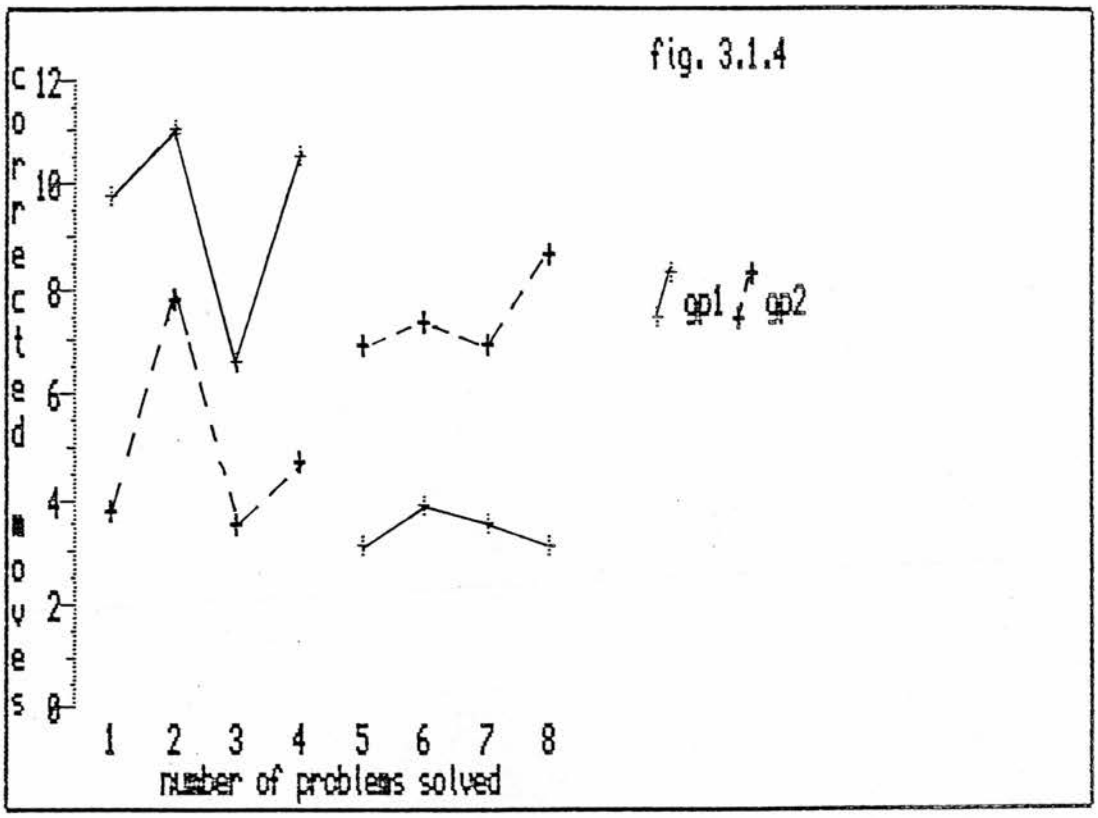
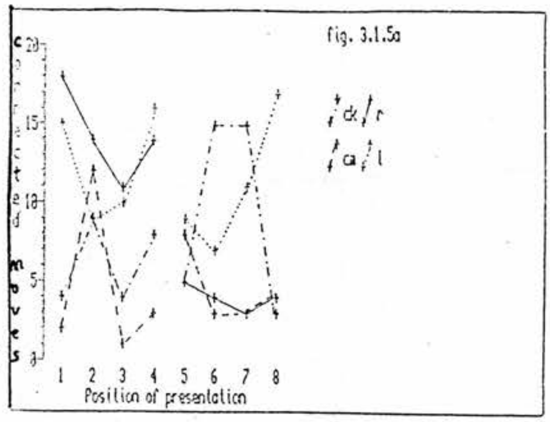


Figure 3.1.4. Corrected moves total as a function of number of problems solved. Problem 5 marks the start of a new problem set for both groups.

Figure 3.1.5 a&b, (overleaf) presents corrected moves data for each of the 8 problems as a function of practice. Once again there is little evidence of transfer and the differences between individual problems is large. An analysis of variance on these data revealed a significant effect of problem ($F_{7,42} = 9.43$; $p < 0.0001$), but no significant effect of practice ($F_{1,6} = 3.43$; $p = 0.1134$), and a problem by practice interaction which just fails to reach significance ($F_{7,42} = 2.17$; $p = 0.0563$). For an analysis of the ismorphic problems in isolation there is a significant effect of problem ($F_{3,18} = 4.23$; $p = 0.0198$) but no effect of practice ($F_{1,6} = 2.98$; $p = 0.1351$), or a position by problem interaction

($F_{3,18} = 2.9$; $p = 0.0581$). Similarly for the non-isomorphic problem considered in isolation there was a significant effect of problem ($F_{3,18} = 8.94$; $p = 0.0008$), but no effect of practice ($F_{1,6} = 2.63$; $p = 0.1559$), or a practice by problem interaction ($F_{3,18} = 0.87$; $p = 0.4739$).

Figure 3.1.5a. Corrected moves total for individual problems of isomorphic problem set as a function of position of presentation of each problem.



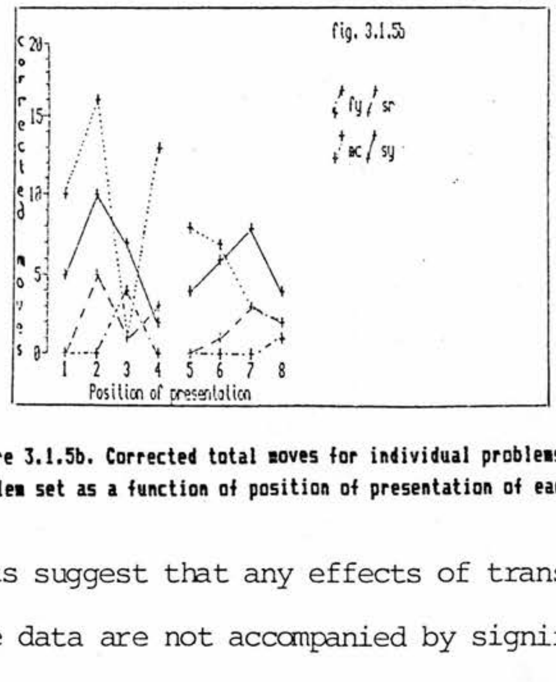


Figure 3.1.5b. Corrected total moves for individual problems of non isomorphic problem set as a function of position of presentation of each problem.

These results suggest that any effects of transfer obtained in the solution time data are not accompanied by significant reductions in the moves to solution. Nevertheless by inspection of figure 3.1.5b, it is clear that once again the MC problem appears to be more difficult than the other problems and does show some slight indication of improvement with practice. What is perhaps most noteworthy is that once again there is no suggestion of transfer within the isomorphic problems and indeed there was a significant effect of problem for this set considered in isolation even though the problems share the same move structure, thus task-structure does not predict problem difficulty. There is certainly no indication for any of the problems that the number of moves to solution approaches a MSP.

Discussion

On the basis of Luger´s approach to problem solving, these state space problems might have been expected to be good candidate problems for a study of domain-specific expertise. The state space analysis provides for a means of defining task domains in a way which Luger considers to have some psychological significance, a position supported by Luger´s own empirical work. However the results of this experiment offer no support for Luger´s hypothesis. The large and approximately monotonic improvement in solution time across problem sets, together with the failure to obtain any corresponding improvement across isomorphic problems testifies to the failure of this hypothesis. These findings suggest that the transfer effects are underwritten not by domain specific expertise as envisaged by Luger, but rather by expertise of a more general sort.

The mechanism responsible for the general transfer obtained here is not easy to discern. It is clear that it is not simply an improvement in some GPS heuristic such as means-end analysis since if this were the case, transfer across all problems including the isomorphic set would havee been expected. But the fact that only the most difficult problems show any transfer and that difficulty is not correlated with structural complexity seems to offer some clue.

The fact that some problems are more difficult than others even when the problems have the same structure (cf the differences in the difficulty observed across the ToH tasks in this study and the significant problem effects of Luger and Bauer), seems good evidence that a major part of the problems solving task lies in the contextual embedding of a problem as distinct from its structural characteristics. In Newell and Simon's terms constructing an appropriate problem space seems to be more difficult for some contextual embeddings than others. This is strongly supported by the work of Hayes and Simon (1976) who, using the ToH task, found that systematic manipulation of the story context of the problems affected not only task difficulty but also the likelihood of transfer between two problems of the same structural type.

One of the most striking features of the two most difficult problems, namely MC and SR, is that subjects attempting to solve these problems are frequently led by the story context into making certain assumptions about the nature of the task that are incompatible with the actual solution path they are required to produce. For example subjects tend not to 'trust' the cannibals to row the boat across the river unattended by missionaries (cf Thomas 1974). Unfortunately there are states in the solution path from which no progress can be made without allowing two cannibals to row the boat across the river. Likewise the subjects tend to represent this problem as a 'ferrying problem' and consequently expect that one person will be in the boat at all times ferrying the others across and that the pattern of moves will be 'two across, one back... and so on'. Both of these assumptions are violated by the actual solution path subjects are required to make.

Similarly for the SR problem it is clear that some subjects view this problem as an 'addition' problem rather than a 'subtraction' problem. That is, most subjects quickly find a means of obtaining 2 tons of sugar simply by emptying the full 5 ton container into the empty 3 ton container. Faced with having 2 tons and needing 4, subjects try and find some means of keeping the 2 tons they already have 'safe' while repeating the process to get another 2 and hence 4. With only the two containers at their disposal this is not possible. The correct solution path requires the subjects to use the 2 tons they have to produce a 1 ton space by partially filling the 3 ton container, filling this space from the full 5 ton container to produce 4 tons directly (5-1 = 4).

While not all subjects come to these problems with such erroneous assumptions it is clear from informal reports taken at the end of testing that some subjects do, (usually those that have the most difficulty), and this might account for the high variance on these problems. This effect of context might be termed contextual error-producing factors, and has much in common with the suggestion of Schulz (1963) that much of problem solving can be usefully regarded as the negative transfer of previously valid hypotheses. It may be that subjects solving problems with these error factors become aware of the need to avoid making such ´common-sense´ assumptions, and it may be that avoiding such reliance on context forms the basis of the general transfer effects obtained here.

One possible means of testing these conjectures would be to examine individual move latencies on these problems, predicting that the moves which required a violation of these expectations would take the most time to be carried out, and that move latencies on these critical states would reduce as a function of the number of such states solved. But as Greeno (1974) has shown, subjects do not solve these problems in a move- by- move fashion, rather they use a good deal of look-ahead, contemplating possible moves before they are reached. Thus it would not be possible to interpret move latencies as reflecting the difficulties associated with the actual move being carried out.

In the absence of such a possibility one prediction that could at least be made is that difficult problems containing these error-producing contexts will produce the most positive transfer to other difficult problems with similar error producing contexts and some experiments associated with this hypothesis are reported in appendix B along with replications of the basic findings reported here and in the following experiment.

Another possible means of addressing this issue and one that finesses the problems mentioned above is to take the simple expedient of asking subjects to make verbal protocol reports explaining why they make the moves they do and why they reject others. There has been much research using protocol reports to allow the experimenter to identify moves which a subject is considering, but there has been little effort dedicated to asking subjects why they carry out the moves they do. This will be investigated as part of the next experiment.

While these suggestions perhaps offers some account of the general transfer obtained in this experiment, it will not in any simple way account for the failure to replicate the isomorphic transfer obtained by Luger and Bauer. If anything, it might have been thought that the benefits of overcoming the counter-productive effects of context would have been greatest in the case of problems with identical structures. Clearly we must look further for an explanation of this failure.

Luger argues that one good index of structurally mediated problem solving is the occurrence of minimum path solutions, ie. solutions in which the goal state is reached from the start state in the minimum number of moves possible. It is clear from the moves analyses presented here that subjects solving the isomorphic problem set in this experiment were not converging on minimum solution paths. On the other hand subjects solving the MC and SR problem did show solution paths very close to the minimums. This is not surprising since a consideration of the state spaces for these two problems (cf appendix A) reveals that if a subject is to provide a solution at all, then such a solution must be in the minimum number of moves possible. There are no non-minimal solution paths available to subjects in these problems, there are only forward and backward moves along this minimum path. In contrast, there are many non-minimum paths for the ToH problem.

On this basis it might be argued that subjects solving the MC and SR problems are forced to search out minimum paths, but they are not required to do so for the ToH problems. Thus subjects solving the isomorphs may have simply failed to utilise structural invariances in order to improve their solutions. Instead they may have been content to produce any path provided it was a solution and required a minimum of cognitive effort. If this were so then the reason for the failure to obtain isomorphic transfer is precisely because subjects are not required to actually use structural invariances in order to build up some domain specific expertise.

On a reconsideration of Luger and Bauer´s results it was discovered that there was a procedural difference between their study and the one reported here, and one that is pertinent to the above explanation. It was found that Luger and Bauer had not simply asked their subjects to solve each of the two problems once. They had in fact used a criterion of mastery procedure. Subjects in their experiment were given one problem which they were required to solve repeatedly until they had produced a minimum solution path in less than 90 seconds. Only when they had achieved this criterion were they allowed to continue on to the next problem which they again solved to criterion.

Clearly the use of such a criterion forces the subject to search for solutions based on the structural characteristics of the problem, in a way in which the subjects in this experiment were not required to do. It may be then that the use of such criteria, requiring as it does that the subject produces an expert solution on each problem before going on to subsequent problems, is a necessary pre-requisite for the occurrence of isomorphic transfer as obtained by Luger and Bauer.

In experiment 2, performance of subjects under this criterion is compared with that of subjects solving problems under a shallow exposure condition of the sort used in this experiment.

Experiment 2: Minimum Solution Paths and Isomorphic Transfer

In this experiment, two groups of subjects solve six ToH problems. One group solves each of the problems to a minimum solution path criterion, the other group is merely required to solve each problem twice. Both groups then go on to solve two non-isomorphic problems, the MC and the SR problem. If indeed it is the case that the failure to obtain isomorphic transfer in experiment 1 was because subjects in that experiment were not required to use structurally mediated problem solving methods then, in this experiment, we would expect to find significant isomorphic transfer for the criterion group only.

It has been suggested that the development of this form of domain specific expertise leads to a concomitant loss in expertise of the more general sort obtained in experiment 1. This view has been put forward by Anzai and Simon (1979) for example, and it can be seen as stemming directly from the trade-off between generality and power highlighted by GPS (see chapter 2). This of course may not be an inevitable consequence of growing expertise for humaan subjects. It could be for example, that if subjects ´en route´ to developing domain specific strategies, also gain some expertise in more general procedures such as ´means-end analysis´then some residual of this expertise might remain for use on other unrelated problems.

The two non-isomorphic problems are included in this experiment as an attempt to assess the extent to which domain specific expertise pre-empts transfer of a more general sort. But the inclusion of such non-isomorphic tasks is only a rather crude measure. If subjects show both isomorphic and non-isomorphic transfer it will be necessary to have a means of factoring out domain specific and more general components of the expertise. In an attempt to obtain a richer source of information about the expertise factors involved here, the method of protocol reporting is used in this experiment in addition to the more conventional measures of solution time and moves to solution.

Protocol reporting has been used with these kind of problems in order to obtain information about the kinds of moves that subjects consider making in their attempts to solve these problems. It was this kind of protocol data that Newell and Simon used to compare human problem solvers with GPS (see chapter 2). In this experiment however, the use of protocol reporting is slightly more ambitious. Instead of merely asking subjects to 'think aloud' about the moves they make, rather subjects in this experiment are asked to introspect about the problem solving strategies they are using, how they are thinking about the problems, whether they recognise subproblem invariances, whether they recognise any similarities between the problems and so on.

By using protocol techniques in this way it is hoped that a more detailed profile of expertise will emerge and one that not only distinguishes between domain specific expertise and expertise of a more general sort but also provides some information as to the psychological mechanisms involved in such expertise.

Method

Problems

The 4 ToH problems used in experiment 1 are again used here and in addition 2 new versions of the problem, the 'harbour' problem and the 'art galleries' problem (see appendix A) to make 6 versions of the 4-ring ToH problem in all. Along with these problems all subjects solved the two difficult non-isomorphic problems from experiment 1, namely the MC problem and the SR problem.

Subjects

The subjects in this experiment were 23 undergraduates of Edinburgh University of both sexes attending 1st and 2nd year psychology lectures. Subjects were not paid for their services.

Design

The 6 ToH problems were presented to all subjects in a counterbalanced order as in experiment 1. Thus all 6 problems were solved by some subjects in each of the 6 possible positions of presentation. This counterbalancing meant that for each group of subjects there were 6 subgroups solving the problems in different orders of presentation. It was hoped that 2 subjects from each group would make up each subgroup, but owing to illness 1 subject is missing from the shallow group solving the problems in the order 'galleries, cakes, rockets, lots, computers, harbours.' Subjects then went on to solve the 2 non-isomorpic problems in the fixed order of MC followed by SR. These were not counterbalanced since the between-group comparison on these problems was considered to be of primary importance.

Subjects were divided into two groups, one group (MSP, N= 12) solved all problems to a MSP criterion which required them to produce a minimum solution path to each problem they solved before moving on to their next problem, unlike the Luger study there was no time limit imposed upon the production of such a solution. The second group were simply required to solve each problem twice before moving on to the next problem; they were not required to produce a MSP or to improve their performance on their second attempt, merely to repeat the problem. These two groups were divided into two subgroups, one subgroup were required to give protocol reports about their solution method the other subgroup were not.

Procedure

As in experiment 1 subjects were told that they would be required to solve a few simple problems "each in the form of a board game". They were told that their solution times and their moves-to-solution would be recorded but that they were to solve the problem only as quickly as was concomitant with not making needless errors. Neither group were told that they would be required to solve the problems more than once, nor were they told how many problems they would be required to solve. Protocol subjects were however told that from time to time they would be required to think about how they were solving the problem and write their thoughts down on a piece of paper. Reports were required after subjects had solved their 1st, 3rd 6th, and 7th problems. On these occasions the subjects were asked to think about their problem solving method and in particular they were asked to consider what parts of the problem they found difficult and why, and what they had learned about the problem that allowed them to overcome these difficulties, and whether they could formulate any useful hints that would help other people solve the problems more easily. They were not asked to recall specific moves nor were they informed of the structural relationships between problems.

After their first attempt at each of the 8 problems subjects in the shallow group were asked to solve the problem again, subjects in the MSP group, provided they had not produced a minimum solution path, were asked to solve the problem again, this time trying to reduce the number of moves they made. They were asked to do this each time they failed to produce a solution in the minimum number of moves possible (15 for the ToH problem, 11 for the MC problem and 6 for the SR problem). As in experiment 1 the experimenter recorded the time taken on each attempt at each of the 8 problems as well as the number of moves to solution. Each experimental session lasted approximately 1 hour.

Results

Isomorphic Problems

Table 3.2.1 presents the mean time, moves and number of attempts at solution over all 6 ToH problems solved by the two groups. This table makes it clear that the two groups of subjects differed very little in terms of the overall level of experience that they received with the ToH problems.

	time	moves	attempts
Shallow Group	1078	270	12
MSP Group	1225	261	12

Table 3.2.1 Comparison between groups of total time (secs), moves, and attempts to solve all 6 ToH problems

T-tests on these data revealed no significant group differences on any measure (time: $P>0.05$, moves: $P>0.05$, attempts: $P>0.05$,). But despite the overall similarity in the level of experience that these two groups recieved on the ToH problems, the transfer profiles show striking differences. Figure 3.2.1 (overleaf) presents the mean total solution time for subjects solving the 6 ToH problems. As can be seen there appears to be a large reduction in solution time as a function of practice for both groups.

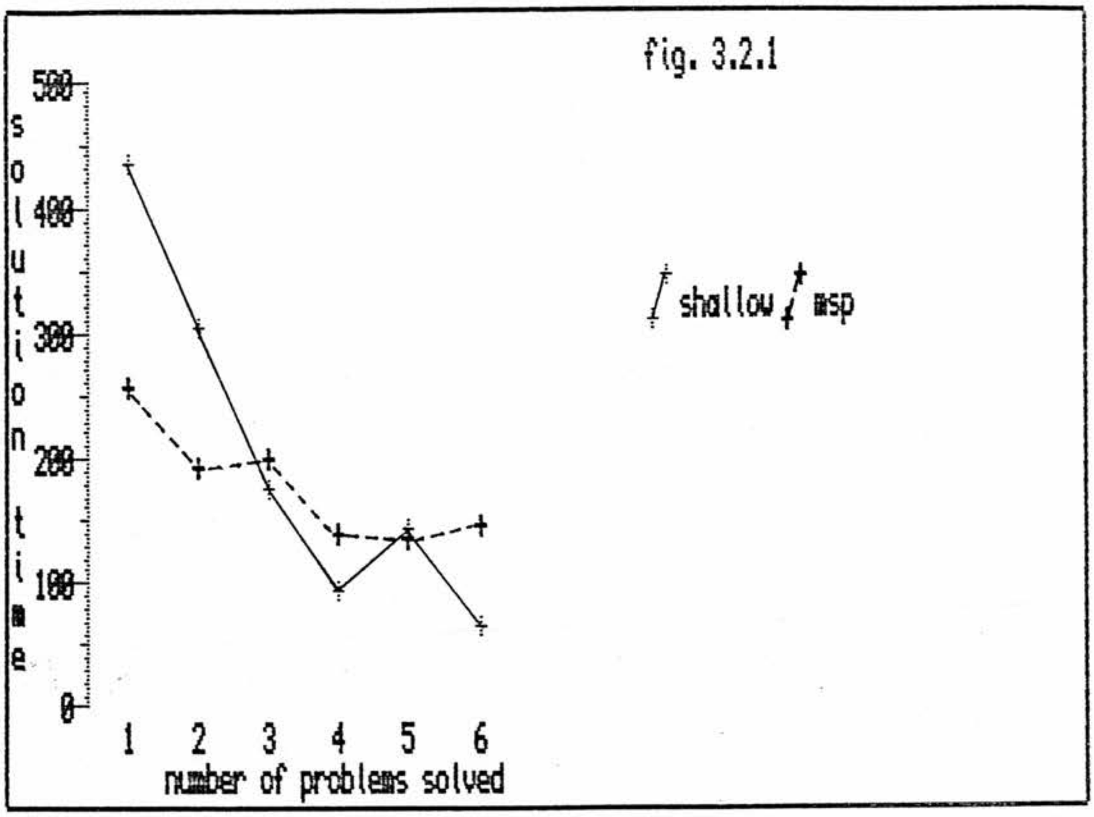


Figure 3.2.1 Group comparisons of total solution time (secs) as a function of number of problems solved for the 6 ToH problems: ALL ATTEMPTS at each problem.

An analysis of variance (BMDP P2V) reveals a significant effect of practice ($F_{5,105} = 8.44$; $p < 0.0001$), but no significant effect of group ($F_{1,21} = 0.30$; $p = 0.05899$). There is however a significant group by practice interaction ($F_{5,105} = 2.54$; $p = 0.0329$). Separate analyses of variance for the two groups however revealed a significant effect of practice for the MSP group ($F_{5,55} = 6.3$; $p = 0.0001$) and the shallow group ($F_{5,50} = 3.5$; $p = 0.0080$). In this experiment then in contrast to the results of experiment 1, both groups show a significant reduction in solution time over isomorphic problems and indeed the shallow group appear to show a larger improvement than the msp group. However when only the first attempt at each of the problems is considered the profile appears somewhat different as is shown in figure 3.2.2.

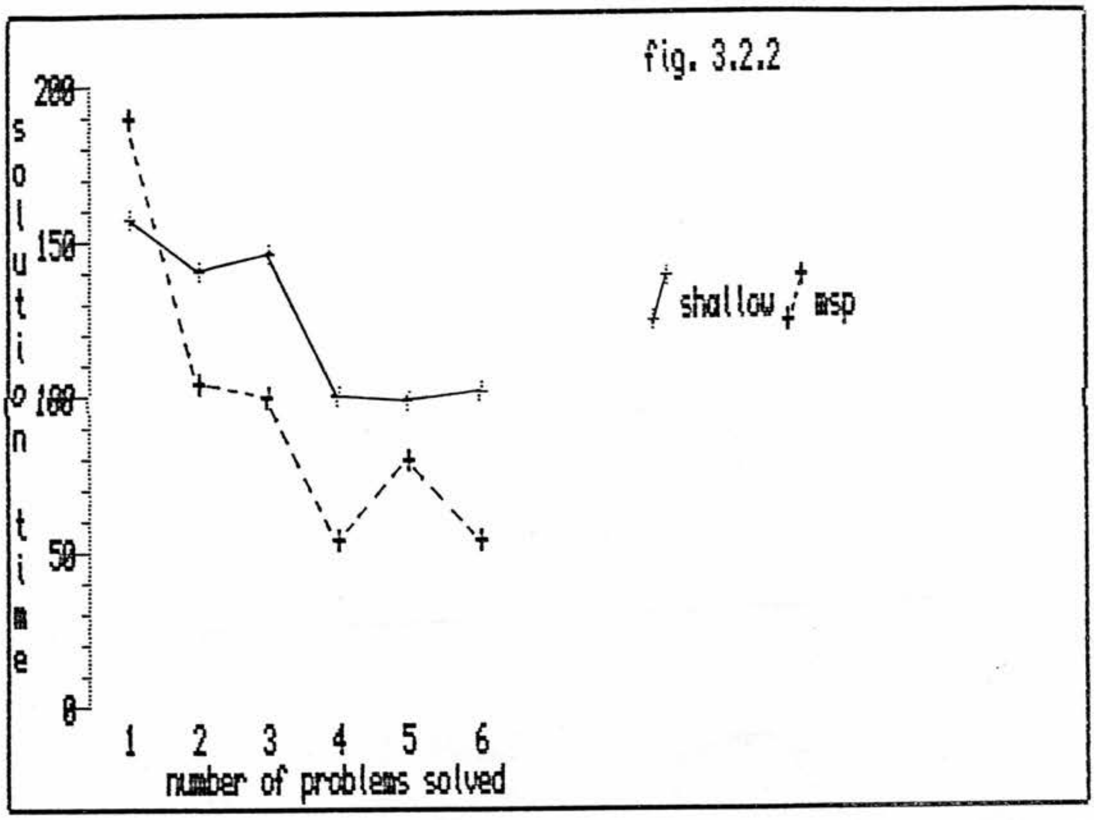


Figure 3.2.2 Group comparisons of solution time as a function of number of problems solved for the 6 ToH problems: FIRST ATTEMPT ONLY at each problem.

Here the MSP group show a greater improvement in solution time than the shallow group. This suggests that the improvement in solution time for the shallow group is largely attributable to intra-problem transfer while the improvement on the MSP group is largely attributable to inter- problem transfer. An analysis of variance on these data while showing a significant effect of practice ($F_{5,105}$ = 4.41; p = 0.0011), did not unfortunately show any effect of group ($F_{1,21}$ = 0.48; p = 0.4979) or any group by practice effects ($F_{5,105}$ = 0.53; p = 0.75). Despite these non-significant group effects on solution time the profile for the moves -to-solution data support this interpretation.

Figure 3.2.3 presents total moves -to- solution data as a function of practice on the 6 ToH problems.

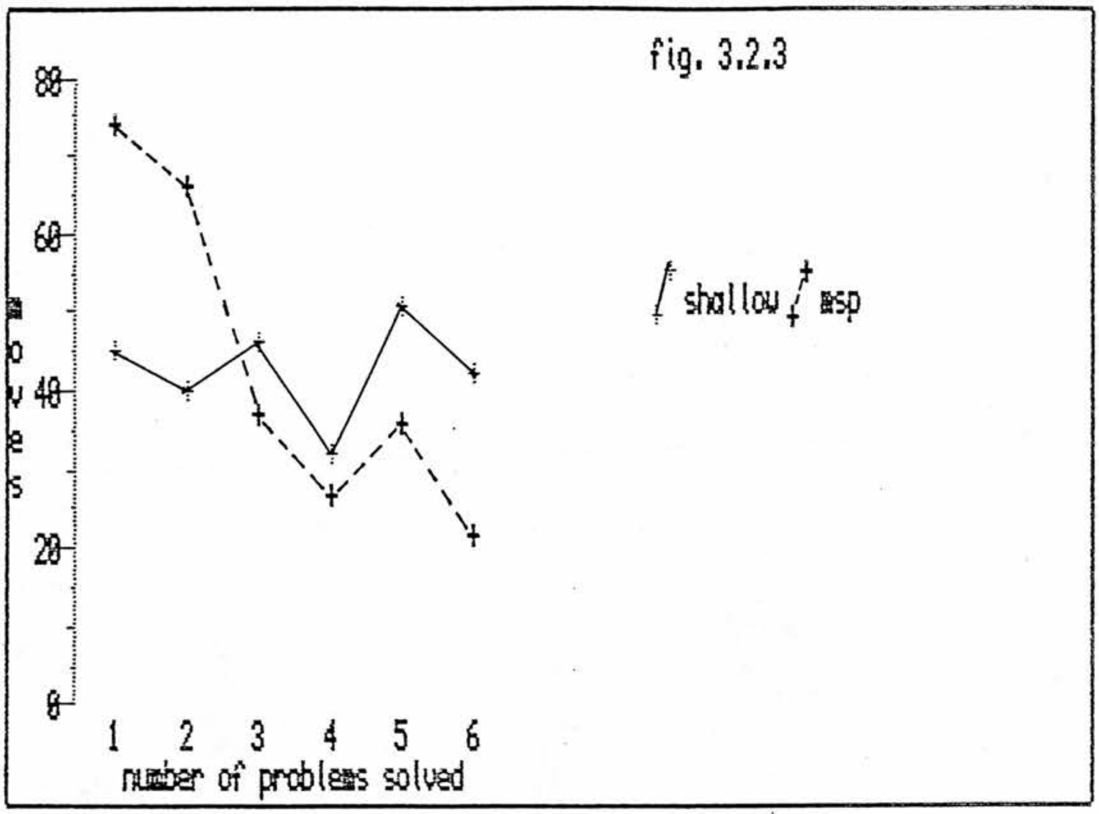


Figure 3.2.3 Group comparisons of total moves as a function of number of problems solved for the 6 ToH problems: ALL ATTEMPTS at each problem.

Here the profile is quite different for the two groups. For the shallow group there appears to be no systematic improvement with practice, while for the MSP group there is a gradual reduction in the number of moves to solution. An analysis of variance on these data reveal a significant effect of practice ($F_{5,105} = 4.41$; $p = 0.0011$), no significant effect of group ($F_{1,21} = 0.05$; $p = 0.8$), but a significant group by practice interaction ($F_{5,105} = 4.2$; $p = 0.002$). Separate analyses of variance on the two groups revealed a significant effect of practice for the MSP group ($F_{5,55} = 5.8$; $p = 0.0002$), but not for the shallow group ($F_{5,50} = 0.66$;

p = 0.65). These results suggest that while the shallow group show an improvement in solution time, in contrast to the MSP group, this is not underwritten by any reduction in the number of moves to solution.

An analysis of the moves data in terms of the percentage of subjects reaching the MSP criteria as a function of practice further refines this analysis. This data is presented as figure 3.2.4.

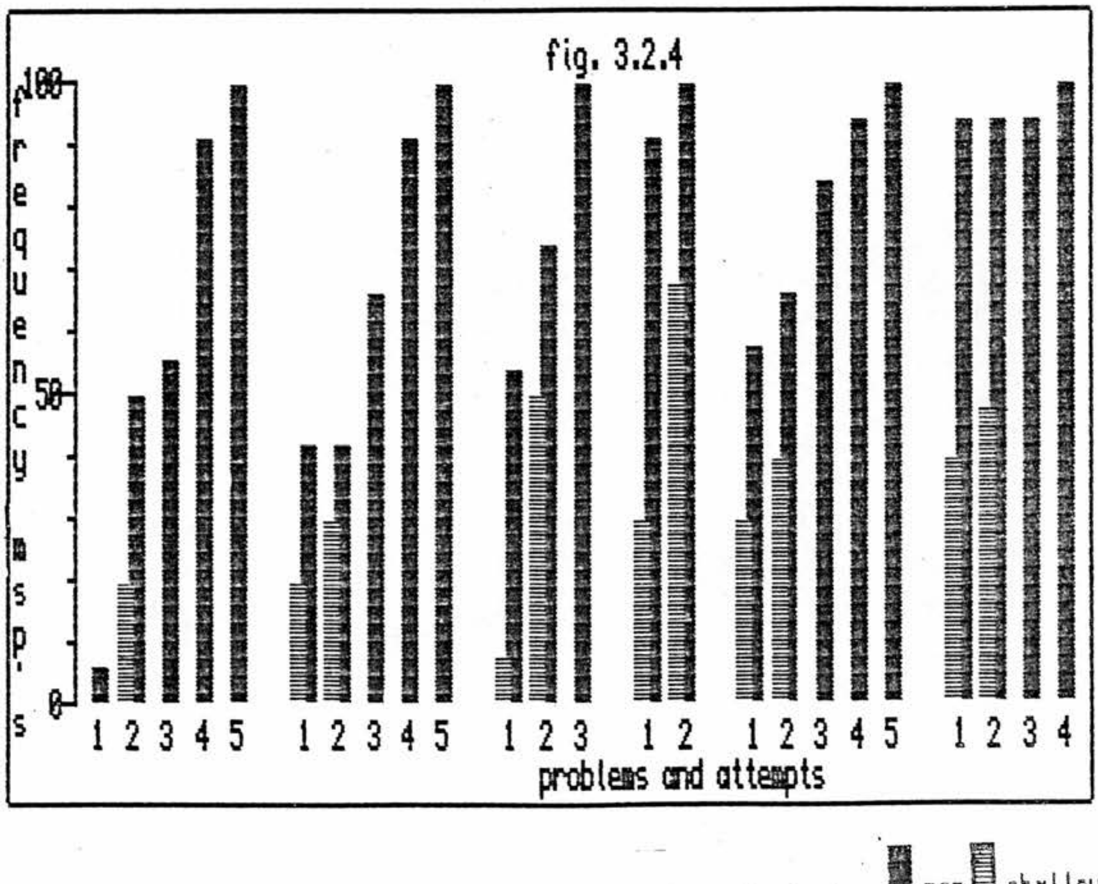


Figure 3.2.4. Group comparisons of the Percentage of subjects producing MSP solutions as a function of the number of attempts at each of the 6 ToH problems.

As can be seen by inspection, once again there are striking differences between the two groups both in terms of intra- and inter- problem effects. The data for the MSP group indicate that on their early problems subjects took many attempts to reach criterion, with most subjects taking 4 attempts or more to reach criterion on their very first problem. But by the 6th problem over 90% of subjects solved this novel problem to criterion on their first attempt (In fact only 1 subject failed to do so). Thus for this group there is an obvious shift from intra- problem transfer to inter- problem transfer as a function of practice, to such an extent that the inter- problem component completely abolishes any intra- problem effects after five problems have been solved.

In contrast the shallow group show little evidence of the development of minimum path solutions. There is some indication of an intra- problem effect with a higher incidence of MSP solution on the second attempt at problems but even this effect is small with the frequency of MSP solution exceeding 50% on only one problem. There is certainly no indication whatsoever of an inter- problem effect. It would appear then that this group in contrast to the MSP group do not exploit the structural invariances in these problems in any significant way.

It is argued from these results then that despite the fact that subjects in the two groups received the same amount of exposure to the 6 ToH problems the problem solving profiles are quite different and that by being forced to produce a MSP solution subjects in this condition are able to generate a high level of expertise in the solving of these problems. So much so that after five problems nearly all subjects could solve a completely novel problem perfectly on their first attempt. Subjects who do not experience the MSP criteria do not develop such efficient problem solving methods on these problems. If anything subjects in this group appear to find some solution path and then repeat that path more quickly on their second attempt at the problem. They do not appear to show any improvement across problems. The next question is to assess whether the expertise generated in the MSP condition produces a marked decrement in transfer to problems with a different structure in comparison to the shallow group.

Non-isomorphic Transfer

Table 3.2.2 presents the mean number of attempts to solve the two non-isomorphic problems presented after the 6 ToH. Naturally the shallow group made two attempts at each problem but it is striking that despite the difficulty of these problems (see experiment 1) the MSP group produced minimum solution paths to both problems in little over a single attempt. 100% of subjects solved the SR problem in a single attempt and 67% of subjects solved the MC problem in a single attempt. In contrast only 50% of the shallow-group subjects produced a minimum solution path on the SR problem on either attempt and only 20% did so for the MC problem. When solution time and moves to solution are considered for the two groups on only their first attempt at the problems there appears to be little difference in the performance of the two groups as is indicated by table 3.2.3.

	MC	SR
Shallow Group	2.0	2.0
MSP Group	1.3	1.0

Table 3.2.2 Group comparisons of the number of attempts taken to solve the 2 non-isomorphic problems.

	MC		SR	
	moves	time	moves	time
Shallow Group	18.2	350	7.6	289
MSP Group	18.9	387	7.5	279

Table 3.2.3. Group comparisons of the moves and time (secs) to solution for the 2 Non-isomorphic problems: FIRST ATTEMPT ONLY.

An analysis of variance on these data revealed no significant effect of group for either moves ($F_{1,20}$ = 0.03;p =0.87), or time ($F_{1,20}$ = 0.02; p = 0.68).

The results of the isomorphic transfer phase suggest that the high degree of expertise generated by the MSP group on the ToH problems does not produce a loss in performance on subsequent non-isomorphic problems at least in comparison to subjects who receive shallow exposure to the same problems. Indeed if anything subjects in the MSP group show a more efficient solution to these non-isomorphic problems if the incidence of minimum solution paths is taken as an index of efficiency. Furthermore this higher incidence of minimum solution paths in the MSP groups does not appear to incur penalties in the form of longer solution times as might have been expected.

Protocol Reports

Given the moves and time analysis of the previous section it is of interest to analyse the protocol reports for a number of key featúres. Given the higher incidence of MSP solutions and the inter- problem transfer of the MSP group then it should be that this group will show a higher incidence of protocol reports that confirm that subjects have recognised the similarity between the isomorphic problems and also that they are attempting to solve the problems in the minimum number of moves. This structural awareness might also be evidenced in the type of strategies that subjects in this group report. For example MSP solutions rely on an exploitation of the recursive sub-structure of the ToH problems and consequently may be evidenced by reports of the type referred to by Simon (1975) such as recognising the need to " get the big disc from the start peg to the goal peg, which means getting the other three pegs discs to the middle peg" and so forth. These kind of reports might contrast with subjects who report using trial and error for example. It is also of interest to ascertain whether subjects are sensitive to particular difficult states in these problems. This kind of analysis can be done both for the reports as a whole and in terms of the protocol reports taken after varying amounts of practice. Table 3.2.4 presents the percentage of protocol reports falling into these various categories.[1]

	recognise structural relations		strategy report		recognise difficult points in sol. path	recognise procedural constraints	others
	sim. of iso's	diss. of non-iso's	struc. mediated strategy	trial & error			
SHALL.	80%	20%	50%	80%	0	0	40%
MSP	100%	66%	85%	35%	66%	0	33%

Table 3.2.4. Percentage of protocol reports falling into the five categories.

By and large the differences are not great, particularly when it is noted that the total number of subjects providing reports was only 11. Both groups appear to recognise similarities between problems but a much smaller percentage of subjects in the shallow group actually report using such strategies. Of note also is that there was a much higher incidence of trial and error reports from the shallow group, and a much higher incidence of recognising difficult points of certain problems in the MSP condition. Such reports took the form of statements that suggested that the layout of certain of the ToH boards made solutions more difficult, and that for the MC problem, the requirement to bring passengers back made the problem more difficult. The category termed ′other strategies′ refers to such reports as "the thing to remember is plan ahead".

[1] See appendix D for details of the categorisation.

While such an analysis of the protocol reports is gross, the differences between the groups is in the expected direction, with the MSP group showing a greater awareness of stuctural factors and less frequent use of non-directed strategies such as trial and error. However it is striking that a high percentage of shallow group subjects show some awareness of the similarities between problems. It is disappointing however, in the light of the moves analysis that there was not some indication in the protocols of subjects' sensitivity to the procedural constraints under which they were working.

Table 3.2.5 presents the same categories as a function of practice. But given the somewhat disappointing global analysis above it is perhaps unsurprising that table 3.2.5 is unrevealing.

	recognise structural relations		strategy report		recognise difficult points in sol. path	recognise procedural constraints	others
	sim. of iso's	diss. of non-iso's	struc. mediated strategy	trial & error			
SHALL.1	--	--	20%	80%	0	0	0
SHALL.2	40%	--	40%	20%	0	0	40%
SHALL.3	60%	--	20%	20%	0	0	0
SHALL.4	--	0	0	40%	0	0	0
MSP 1	--	--	60%	20%	0	0	20%
MSP 2	80%	--	60%	0	40%	0	0
MSP 3	80%	--	80%	0	20%	0	0
MSP 4	--	60%	20%	20%	0	0	0

Table 3.2.5. Percentage of protocol reports falling into the five categories as a function of practice.

Several points however are worth noting. Firstly there is a high incidence of trial and error reports particularly in the early protocols. It is also worth pointing out here that subjects tend to report trial and error in response to their first rather than subsequent attempts on a given problem. For the MSP group it is clear that reports of structure mediated strategies appear only after some experience with the problems despite the fact that subjects appear to recognise similarities between the problems early on.

This recognition of similarity did not occur quite so early for the shallow group subjects. Finally for both groups although there is a move away from trial and error solutions over the ToH problems occurrence of the non-isomorphic MC problem immediately preceding report number 4 seems to have produced a tendency to revert to trial and error especially for the shallow group.

These reports are broadly consistent with the view that subjects in both groups recognise the structural affinities that exist between the ToH problems but that this happens more readily in the MSP group and this group has a higher tendency to utilise such knowledge in developing structurally mediated strategies.

Discussion

The results of experiment 2 confirm the hypothesis that MSP criteria are necessary in order to obtain the structurally mediated transfer effects postulated by Luger and Bauer. While shallow exposure to these tasks can, under certain circumstances, produce an improvement in solution time across isomorphic problems, such improvement is not accompanied by a shift towards optimal solutions based on the exploitation of invariance. This suggests that a high degree of expertise within a single problem is necessary for such isomorphic transfer to occur. The fact that both shallow and MSP subjects in this experiment received the same amount of overall exposure to the 6 problems suggests that such expertise is not just attributable to a greater exposure to the problems. Rather such expertise is a response to experimental procedures requiring a search for optimal solution paths.

It must be noted however that even with such stringent procedures as the MSP criterion in operation, changes in the surface structure form of the isomorphic problems still presents some difficulty for subjects in the early stages of learning. It is clear from the comparison of intra- and inter- problem effects in this condition, that subjects must not only learn appropriate structurally mediated solutions, but that they must also learn to recognise the problem domain for which such methods are effective. This implies that high intra-problem learning in the early stages followed by inter-problem experience in the later stages would be the most efficient procedure for generating the isomorphic transfer obtained here. This conclusion is consistent with the learning set work of Harlow and others reported in chapter one.

One factor of importance to emerge from the protocol analysis in this experiment is that subjects in both groups seem to recognise similarities between isomorphic problems but only the MSP group go on to exploit such similarities in their search for optimal solutions. This difference may be attributable to the fact that although the shallow group are given enough inter- problem experience to be able to recognise such similarities, they are not given enough intra- problem experience to be able to develop effective procedures for producing minimum solution paths. An alternative explanation and the one preferred here, is that subjects in the shallow group have the option of developing such solutions but the constraints that the shallow procedure imposes on these subjects does not make it a sensible option for them. It is certainly the case that the differences between the two groups seem to reflect the differences in the procedural constraints they experienced.

The MSP group were required to solve each problem repeatedly until they produced a minimum solution path, thus it was to their advantage to find, as quickly as possible, a solution method that would allow them to exploit similarities between problems and thus reduce the number of attempts they made at each problem. In contrast the shallow group were required to solve each problem twice irrespective of whether or not they produced a minimum solution path on their first attempt. Thus it was not to their advantage to invest resources in searching for optimal solutions, they would do well simply to find any solution as quickly as possible on their first attempt and then to repeat it as quickly as possible on their second attempt.

If this view is correct, then it indicates that subjects are not only sensitive to problem structure but also to the functional requirements that the experimental procedure imposes upon them. Subjects not only learn about structural invariances present in a problem domain but also functional invariances, namely, what it is that the experimenter requires them to do with the problems. Thus contrary to Luger´s approach it is inadequate to define a problem solving task solely in terms of its structural characteristics. Any theory of problem solving based on such an assumption is an oversimplification. In contrast to the expertise indexed by isomorphic transfer, this sensitivity to functional goals is presumably an aspect of expertise of a more general sort. It may be this more general component of expertise that is responsible for the performance fo the MSP group on the non-isomorphic problems. Here their sensitivity to the need to produce a MSP solution led them to keep restarting the problem from the beginning as soon as they recognised they had strayed from the minimum solution path and before they had actually solved the problem from beginning to end. An interesting question that arises from this view is what kind of problem solving behaviour would have been observed if subjects had been fully informed as to the procedural criteria in operation in this experiment? It may be for example, that a major 'problem' for subjects is in fact deciding what procedural criteria are in operation and thus what problem solving methods would be appropriate. If subjects had been informed in this way then quite different profiles might have emerged. For example, a subject, told he was to solve 6 structurally identical problems each to a MSP criterion, might invest a great deal of effort in finding an MSP solution on his first problem, he would probably consider each move very carefully avoiding moves he thought would take him off the minimum path. He might even avoid making any moves at all until he had worked out the minimum solution path in his head.

At present, the subjects in this experiment could be considered as being in a situation similar to subjects in animal learning experiments wherein extensive pre-training is used to enable the experimenter to signal what it is he wants the animal to do. As Gagne (1964) points out, given the apparent importance of functional criteria, it is surprising that there have not been more studies that have manipulated procedural variables such as the kind of instructions given to subjects and so on.

Taken together, these experiments have demonstrated the usefulness of multiple-problem transfer paradigms as a means of identifying expertise factors in problem solving. But what the results imply about the psychological mechanisms involved in such expertise is less clear. This will be taken up in the general discussion.

General Discussion of the State Space Experiments

These two experiments demonstrate that it is possible to obtain quite reliable transfer effects within a laboratory setting provided that proper experimental procedures are used. All of the main effects reported here have been replicated and are reported in appendix B. The experiments have used transfer effects to distinguish empirically between the two forms of expertise discussed in the review chapters. They provide evidence both for the more general form of expertise and the domain specific sort, and show how the kind of transfer generated depends on the kinds of experimental procedure used. The general expertise factor tapped by the procedures used in experiment 1, produced only a weak form of transfer when compared with the domain specific expertise generated by the procedures of experiment 2. Indeed this expertise produced 100% transfer after 5 problems had been solved. This dramatic transfer effect, highlighting as it does both the intra- and inter-problem components of expertise, provides the basis for a working definition of domain specific expertise, namely; that which renders formally equivalent problems psychologically equivalent.

The fact that the kind of transfer observed depends largely on the kind of experimental procedures used is an important one. This fact explains why previous research (Luger and Bauer, Reed et al, Thomas op cit) produced such inconsistent results. More importantly, it is possible to see that because of this fact, the theories developed on the basis of this research are limited in their explanatory power to phenomena generated by that research. In particular, the fact that these other studies concentrated solely on the train-and-test transfer paradigm, means that their explanations emphasise the intra-problem aspects of expertise at the expense of the inter-problem components.

For example, Luger's theoretical emphasis is to explain the performance on a single ToH problem in terms of the acquisition of knowledge about sub-problem invariances as formalised in the state space graph. Indeed, by looking in detail at subjects' move patterns on repeated attempts at a single ToH problem, he has shown that their move-patterns come to mirror the sub-problem invariances as their experience with the problem grows (Luger and Steen 1981). Luger uses transfer only as a diagnostic tool, to show that the subjects' problem representation is at a sufficiently abstract level to encompass knowldege about structural invariance. He does not use transfer to examine the difficulties presented to his subjects by changes in the surface structure forms of the problems, and he is not concerned to explain why one version of the problem he used is more difficult than the other.

In contrast, the multiple-problem studies reported here, are chiefly concerned with transfer as a psychological problem; that is to say, they seek to explain why changes in surface structure form are an obstacle to the transfer of well-learnt problem solutions. As such, these studies have shown that the intra-problem learning of the sort explained by Luger, is only one part of a full account of expertise.

They have shown that the attainment of mastery on a single problem does not guarantee mastery of a second problem, even when that second problem is formally equivalent, solvable by the same strategy, and when the subject has recognised that the problems are the same. Only when a subject has had a large amount of exposure to different surface forms of the same problem, in addition to demonstrating mastery within a single problem, is he able to show 100% transfer. In short, there is a significant inter-problem component which has to be explained by a theory of expertise. Such a theory would not be in opposition to, or in replace of, Luger´s account of intra-problem learning but in addition to, and building onto, this account.

By considering the inter-problem component of expertise as a process of ´translation´, its relation to the intra-problem component can easily be appreciated; Taking Luger´s account of intra-problem learning, then reaching a MSP criteria within a single problem implies that the subject has some problem space which allows him to exploit the sub-problem invariances of the ToH structure. On being presented with a second problem, he must find some means of encoding the new text instructions into the same problem space. Once he has achieved this, then he will be in a position to take advantage of the search strategies he has already developed for that space (Anzai and Simon 1979). If a subject shows transfer between the two problems, it is by virtue of him having the ability to make this translation. If a subject fails, either he does not have translation procedures required or he does not recognise the problems as inter-translatable. Since in our experiment, subjects report recognising the similarities between problems even when they fail to show transfer, we must assume the failure is one of translation and not recognition. Thus, the subject's growth in expertise, is explained by gains in his ability to translate different problems into a single underlying problem space.

From this outline, it is clear that the burden of explanation for an understanding of the inter-problem component lies in an explanation of the factors which govern how a subject comes to interpret new versions of the same problem in terms of a single existing problem space, and how this improves with experience.

But while the ToH studies reported here have been valuable in demonstrating the need for a theory of the inter-problem component of transfer, they have yielded little by way of insight into the nature of this translation process or how this might change as the subject's experience grows.

The inability to generate an explanation of the inter-problem component stems, to a large extent, from measurement difficulties associated with the state space problems. In these tasks the major dependant variables are the time and the number of moves taken to move the pieces from their initial state to the goal state. In board games of this sort, time measurements are rather gross, involving as they do, not only thinking time, but also the time taken to carry out the action itself. Measures of the number of moves are also rather gross since for example, some subjects may ´experiment´ with moves, trying possibilities they are not committed to, while others may attempt to work everything out in their heads before making any moves. But there is a more fundamental problem with these variables; these variables are variables relating to the subject´s **search** through the problem space, they are not variables that relate to the subject´s **construction** of a problem space or his **translation** of a new problem into that space. In short, these variables are indices of the intra-problem component of transfer, they tell us little about the inter-problem component.

A third measure that has been used with these tasks is verbal protocol reporting. With the work of Newell and Simon, this has been limited to a recording of the think aloud verbalisations of subjects as they solve a single ToH problem. In effect, this means that the experimenters obtain data about, not only the moves the subjects make, but also the moves they **consider** making but do not make (and of course, the moves they never **even** consider making). Newell and Simon´s work amply demonstrates that such concurrent verbalisations are a rich source data about a subjects search of his problem space, once again however, they do not yield any information about the problems of translation or construction of that space.

In experiment 2, an attempt was made to modify Newell and Simon´s method of protocol reporting in a way that would provide such data. Instead of asking subjects to verbalise the moves they were considering, they were simply asked to provide information about whether they recognised similarities between the problems, and what aspects of the problems they found difficult. In this way it was hoped to find out more about how an everday understanding of the text instructions interfered with the translation process; how for example, certain instructions made it difficult to recognise that previously successful combinations of operators could be re-applied in this new situation. But the results were disappointing: by and large, the reports did not indicate why some problems are more difficult than others, or in any other way elucidate the inter-problem component. The reports did provide evidence that subjects did recognise similarities between problems,but this was seldom expressed in terms of sub-problem invariance. (cf appendix D for example protocols). It seems that these problems are not amenable to this kind of protocol reporting.

These measurement limitations reduce the usefulness of the state space problems as candidates for the further study of inter-problem transfer. But their suitability is reduced even further by a second limitation: The problems do not lend themselves to a systemtatic manipulation of surface structure, and consequently it is difficult to generate large numbers of problems within a single type. As a consideration of appendix A reveals, the generation of new versions of the ToH problems is a rather haphazard affair, and many of the instructions appear quite similar. Certainly the content of the instructions does not vary along any clearly discernable dimension; For example, it is not clear what there is in common between moving lorries around streets, moving wedding cakes from table to table, and changing numbers in a computer memory. Neither is there any principled way of manipulating the sequential structure of the task instructions; for example, it is not clear that presenting information about **what** objects on the board can be moved, before presenting information about **where** they can be moved to, rather than vice versa, is a meaningful aspect of the task instructions. In addition to the rather unconstrained way in which the instructions vary, there is also the fact that each of these problems is solved on a different board, and it seems that changes in the perceptual characteristics of the board contribute in unexpected ways to the difficulty of the different versions of the problem (see appendix D for protocol evidence of this).

In all that has gone before, we have made a fundamental distinction between the process of encoding a problem into a problem space and the process of searching that space for a solution. It may be that underlying the limitations of the ToH problem is the fact that, although the process of encoding is significant for subjects, it is not their primary problem. The state space problems are primarily problems of SEARCH; with these problems, the subject's resources are **not** taken up with understanding the task instructions they are taken up with search of the problem space which **results** from the encoding of task instructions. Consider a typical subject solving the ToH problem: He may spend 4 minutes or so, moving the pieces around the board, searching for a legal combination that gets him from the start to the goal. But he will probably only spend 1 minute reading the instructions. The majority of his problem solving effort is taken up with searching the extant problem space, and physically making moves around the board. In comparison to the problem of search, the problem of encoding is trivial.

For a proper study of the inter-problem components of expertise, what is required is a problem domain in which this asymmetry between encoding and search is reversed; a domain in which ´the problem´ is that of mapping the different surface versions onto a single problem space, a space from which the solution follows trivially. There are in fact several such problem domains in the psychological literature on reasoning, and if we compare the ToH problem with one of these, namely the **transitive reasoning task,** it will be apparent that these kinds of problem offer numerous advantages over the ToH problems for further study of inter-problem learning.

The transitive reasoning task is simply this; a subject is given two or more statements which describe several items along some dimension for example : ´Alan is taller than Bob, Bob is taller than Carl, Carl is taller than Dave´. The subject is then asked a question relating to the premises, such as, ´Who is tallest?´ or ´Is Alan taller than Carl?´. The subject then answers the question by providing a name such as ´Alan´ or by answering yes or no.

The work of DeSoto et al (1961), Huttenlocher (1968), and Trabasso et al (1975), to name just a few, has demonstrated that these problems support a wide range of psychological measures. They have used not only reaction time measures and error rates but also protocol analysis to show that subjects solve these problems by using the premise information to construct some form ´spatial image´

which reflects the relative ordering of the items. Once this ordering is established then subjects can simply 'read off' the answer to the question without the need to make chains of inferences. So for example, the protocol reports might reveal that a subject 'makes a mental picture of the items in a line with the tallest one at the left, shortest at the right' and so on. When asked 'Who is tallest?' the subject simply 'reads the left hand name'. Once the premise information is encoded in this way any number of questions can be answered by recourse to this simple 'look-up' procedure. Such a spatial image device constitutes, what for Newell and Simon would be, the subject's problem space The mapping of premise information into the image constitutes the encoding process, and the process of 'look-up' constitutes the search of the problem space. Using reaction time measures it is possible to distinguish between the process of encoding and the process of search; by measuring premise reading time we obtain an index of encoding, by measuring question answering time we obtain an index of search. An analysis of these different time measures shows that, in comparison to the ToH problem, the relation of encoding to search is reversed; question answering can take as little as 3 seconds compared with premise encoding times of 25 seconds for five-term series problems.

In addition to the sensitivity of the psychological measures, associated with these problems, it is a straightforward matter to generate any number of logically equivalent problems which differ systematically in their surface structure form. For example, it is possible to change not only the number of premises contained in a problem, but also the order of presentation of the premises and the kind of relational term used in the premises. None of these surface structure changes affects the logical equivalence of the problems, yet the research has shown that they affect the time taken to solve these problems; for example; C<B,A>B takes longer to solve than A>B,B>C. This has been taken as evidence that surface structure variation affects the encoding process in significant ways.

Thus, research with these problems, in contrast to our experiments with state space problems, has shown that protocol analysis can be used to examine in some detail, the kind of problem space a subject uses, and also that reading time measures can be partitioned into those that are directly related to the process of encoding and those related to search. But to what extent can we be sure that these problems can be used to generate task-specific expertise of the sort we observed in the ToH problems?

There has been some evidence, and much conjecture, about the role of expertise factors in the solving of these inference problems. The work of Quinton and Fellows (1975) for example showed that there were marked differences in solution time between novice subjects and more experienced subjects solving the 3-term series problem. In a review, Johnson-Laird (1975) conjectured that discrepencies between the findings of several researchers could be explained in terms of the different amounts of practice they gave to their subjects. For example he argued that Huttenlocher's spatial image strategy was a product of novice performance on the task and that more experienced subjects moved away from this strategy to one based on a linguistic (propositional) representation of the premise information of the sort hypothesised by Clark (1969), and finally to a ´perceptual´ (syntactic) strategy which exploited invariances in the way in which premise information was presented (Wood 1978 and Quinton and Fellows 1975). Each of these three methods relies on a qualitatively different way of representing the premise information. This view of expertise was also adopted by Ohlsson (1980), who used the data from these different experiments to build a A.I. model of ´the path-to-expertise´ conjectured by Johnson-Laird. These findings are reviewed in some detail in the chapter that follows, but this brief outline will suffice to highlight an important difference between the view of expertise based on these findings and that developed, in response to our experiments with the state-space problems.

In the case of the state space tasks we argue that the subject´s problem space remains stable over practice, and that inter -problem transfer is obtained by virtue of improvement in the process of encoding into that space. But in the case of the inference problems the situation is different; it would seem that subjects actually **change** the kind of problem space in which they work: If a subject shifts from a spatial analogue representation to a propositional linguistic representation, then clearly he has shifted to a representational device which is qualitativly different. The inter-problem component of transfer is not simply a matter of improvements in encoding processes, it is a matter of quite radical changes in the way of representing the problem domain.

Such changes in representation have important implications for the generality of the expertise generated in these tasks. Quinton and Fellows for example, argue that their expert subjects develop a strategy which is so closely tied to the syntactic invariances in the problems that minor changes in the surface structure would make the strategy ineffective. So, far from being a form expertise that renders formally equivalent problems psychologically equivalent, the reverse seems to be the case, subjects have moved from a strategy which could be extended across the whole range of transitive problems (the image device) to one which solves only a subset of those. This view of expertise as a shift to 'shallow heuristics' is inconsistent with the view developed in this thesis.

Thus there is a tension between the account of expertise developed in the transitivity research and that we have developed with the state space problems. This tension raises a number of empirical issues. Firstly is it possible to demonstrate a form expertise more akin to that which we have observed in the state space studies, and will the affluent psychological measures available in this new domain help to account for the inter-problem component of that expertise? Secondly, how are we to reconcile this view with the view of expertise developed by Johnson-Laird and others? If we are to posit two qualitatively different forms of expertise then what are the factors that govern which kind of expertise will develop? In the chapter that follows, the work with transitivity is reviewed in detail and these empirical issues are assessed in the experimental studies of chapter 5.

CHAPTER FOUR

DOMAIN SPECIFIC EXPERTISE AND THE SOLVING OF TRANSITIVE INFERENCE PROBLEMS

Inference as Problem Solving

Inference problems are characteristic of a problem solving task in as much as they involve goal-directed behaviour. So for example, in the case of transitive inference tasks, the subject is given information in the form of a set of premises (ie. ´John is taller than Paul, Bob is taller than John´). He then has to process this information in response to a specific goal which is provided in the form of a question such as ´Is Bob shorter than Paul?´.

As was briefly outlined in the preceding section, the transitivity problems are a particularly useful class of problems for the study of inter-problem transfer. This is so for several reasons; firstly, we can define the class in terms of the logical equivalence of its members, and it is possible to systematically manipulate the surface form of the problems within this logically defined class. Take for example the two problems below;

Alan is taller than Bob. Bob is taller than Carl.

Is Alan taller than Carl?

Carl is shorter than Bob. Alan is taller than Bob.

Is Carl shorter than Alan?

These two problems are logically equivalent since they both involve the coordination of three items along a single dimension in order to infer which of the three is the furthest along that dimension. Yet they differ markedly in their surface structure forms. This has been achieved by changing the order of presentation of the two premises and the order of presentation of the items within the first premise. A change as also been made to the order of presentation of the items in the question.

By changing the order of presentation of premise information in this way 8 possible surface forms of the 3-term problem can be generated. If 4- or 5- term problems are used then even more versions are produced. In addition, there are 4 versions of the comparative question (A>C? A<C?, C>A?, C<A?) yielding 32 different versions of the problem in all. This family can be further extended varying the question form; Instead of a subject being asked the ´comparative´ form of the question as above he could instead be asked the question in ´superlative´ form, ie. ´Who is tallest/shortest?´, yielding a further 16 basic forms. The nature of the relational terms used in the premises affords a further dimension for extension.

It should be clear then that this problem domain benefits over that of the state-space tasks used in previous experiments with respect to at least one criterion, namely having a large domain of problems which can be systematically manipulated. As the following review of the literature reveals, these problems offer a further advantage; they support a varied range of psychological measures which provide insights into the nature of the representational mechanisms that underwrite performance in this task domain.

Psychological Mechanisms in Transitive Inference

There are major lines of research on transitive inference in both the developmental and adult literature and in both areas there has been much attention paid to one particular psychological mechanism, namely the spatial image device first discussed by DeSoto et al (1965). While there have been many other psychological models (cf Hunter 1957, Clark 1969) put forward to explain performance on these problems, the spatial image device has proved to be by far the most influential and consequently it is discussed below.

Huttenlocher's work on transitive inference (1968) developed the spatial image account first postulated by DeSoto et al. She used the method of verbal protocol reporting to examine how people thought they were solving the problems. These revealed that subjects claim to be constructing a mental picture of the items referred to in the premises. This picture consisted of an ordering of the three items along a horizontal or vertical axis. Some subjects reported using the first-mentioned item in the first premise as the first item in the mental array and construct the array accordingly, while others fix the dimension of the array first (ie. ´tallest goes on the left´) and sort the premises accordingly.

In studies both with children arranging actual blocks in accordance with information presented in premises, and with adults solving the more familiar form of the three-term problem, Huttenlocher found systematic differences in performance according to the particular version of the problem being solved. In particular she found that problems in which the third item was introduced in the second premise as the grammatical subject of that premise were much easier to solve than problems in which it was introduced as the grammatical object. Thus, given the first premise ´A is above B´, if the third item is introduced by the premise ´C is below B´ it is easier than if it is introduced by the premise ´B is above C´.

In explaining these somewhat counter-intuitive results Huttenlocher argues that comprehension requires a correspondence between the linguistic form of the statement and the situations described by that statement. If the statement is incongruent it must be reordered by subjects before it can be integrated into the image. This means that for Huttenlocher the easiest form of the three-term problem is ´A is taller than B, C is shorter than B´, while the most difficult form is ´B is shorter than A, B is taller than C´. These predictions are consistent with the findings of DeSoto et al and were confirmed in Huttenlocher´s own studies.

Developmental research has also made appeal to some form of spatial image device as the mechanism for solving transitivity problems. Trabasso et al (1975) for example have used five- and six-term variations of the three-term series corresponding to problems of the form 'A is taller than B, B is taller than C, C is taller than D, D is taller than E' for example.

In these studies children were repeatedly presented with the premise information until they had memorised each of the 4 pairwise comparisons (ie. A is taller than B etc). They were then asked questions about every possible combination of items (ie. 'Is C taller than B?', 'Is B taller than D?' etc). In this way it was possible to test subjects on large numbers of questions which varied according to the number of transitive steps there were between the two items mentioned in the question. A question such as 'Is A taller than B?' required no inferential steps because the information was actually presented in the premises directly, but a question such as 'Is B taller than D?' requires the subject to make one inferential step via the middle term C, and the question 'Is A taller than E?' requires three inferential steps via B, C, and D.

Using this procedure and measuring question answering time, Trabasso et al obtained the somewhat counter-intuitive result that the <u>more</u> inferential steps a question required the subject to make the <u>faster</u> the subject was able to answer the question. Trabasso et al argued that these data could not be explained by any model that required subjects to actually carry out a chain of transitive inferences since this would predict that more links in the chain would require longer processing time. Instead Trabasso et al argued that this reaction time profile, referred to in the developmental literature as the inference distance effect, is good evidence for the subjects´ use of some form of spatial image representation. If subjects were constructing a linear array in their mind´s eye and simply reading off the answer to the question from this array then naturally, items that were further apart in the array will be more easily discriminated and hence responded to more quickly.

These findings demonstrate that the transitive inference problems offer a broad range of measurement techniques which provide an affluent source of information concerning the psychological mechanisms involved in the solving of these problems. Furthermore it would seem that the spatial image device is a strong candidate for one such psychological mechanism. But what of transfer effects and the role of expertise factors in the solving of these problems? Where does the spatial image device sit with respect to expertise? Is the use of the image device a relatively sophisticated strategy emerging only after extensive practice or is it a rather naive response to the task domain and one that is quickly abandoned in favour of more sophisticated methods? Is it the case then that psychological variables such as the inference distance effect and the effects of problem form diminish as a subject becomes more expert at the solving of these problems?

Expertise Factors in the Solving of Transitivity Tasks

The notion of a path-to-expertise has been used in the adult literature on transitive reasoning to describe the way in which a subject´s performance on three-term series problems might change as a function of practice (Ohlsson 1980). In this view it is suggested that naive subjects solving these problems might use quite general but rather inefficient problem solving methods, but as they become more experienced at solving the problems they develop efficient methods specifically tuned to the task domain under consideration.

Both Johnson-Laird (1975) and Ohlsson (op cit) adopt this view, suggesting that novice subjects begin by the use of a Huttenlocher-type spatial image device but as they improve they move towards using some form of linguistic representation (cf Clark op cit) and finally use a very efficient strategy based on selective search of premises. Evidence for this hypothesis comes form the work of Quinton and Fellows (1975) and Wood (1978).

In the study of Quinton and Fellows subjects were presented with large numbers of three-term problems to solve. Each problem was presented textually and subjects were required to answer one of two possible superlative questions (ie. Who is tallest?/Who is shortest?). Quinton and Fellows not only measured the total time taken for subjects to read the premise and answer the question but also they asked subjects to provide verbal protocols about their solution methods. These were taken at intervals throughout the experimental session. By using this method Quinton and Fellows were able to identify two basic kinds of method which they termed, 'thinking strategies' and 'perceptual strategies'.

The 'thinking strategies' can be further divided into two types 'series formation' and 'elimination'. Subjects using series formation report taking the items mentioned in the premise and forming a series or linear array in their mind's eye from which they then read off the answer to the question. Elimination on the other hand, does not involve the formation of an integrated spatial image, rather subjects using this method first read the question and then search the premises for the relevant item. Thus if the question asked for the tallest item and the first premise read 'A is taller than B', the subject would retain only item A in memory as the tallest item currently encountered and go on to the next premise. If the second premise was 'B is taller than C' then he would have his choice of A confirmed. If on the other hand the question had been 'Who is shortest?', then the subject would have eliminated A has a possible candidate on his reading of the first premise and eliminated B as a candidate on his reading of the second premise, leaving C as the shortest item.

The perceptual strategies identified by Quinton and Fellows were a more sophisticated version of elimination which allowed the subject to avoid considering the meaning of the premises altogether. In their most advanced form they involve the following operations:-

1) The subject would compare the relational term used in the question with that used in the first premise. If they matched, the item on the lefthand side of the premise was carried over as the candidate answer. If they didn´t match, the other item was carried over. Thus for the question ´Who is shortest?´ and the premise ´A is taller than B´, the terms do not match and so B is carried over.

2) Next, the subject searches the second premise for an occurrence of the candidate term, if the term does not appear then it is the answer to the question, if it does occur anywhere in the second premise (ie. it is in fact the middle term of the three) then the other term in that premise is the answer.

Quinton and Fellows found that subjects solving large numbers of these problems gradually improve their performance as assessed by solution time. They also found that the type of strategy that subjects used varied as a function of practice. Novice subjects tended to use some form of thinking strategy while more practiced subjects tended to use some form of perceptual strategy. Thus they argued that the transfer effects they obtained were underwritten by a qualititive shift from thinking strategies to the more perceptual strategies.

They also argued that one of the characteristics of this path-to-expertise was a trade-off between generality and power. They suggest that although thinking strategies employing the image device would be comparatively slow they would in principle enable the subject to deal with any form of question or premise he might be presented with. Even negated premises and comparative questions are solvable in this way. In contrast while the perceptual strategies are highly efficient methods they are quite task specific in as much as they could only be used to answer superlative questions and were limited to regular forms of the premises.

This view of a path-to-expertise as a trade-off between generality and power is supported by the work of Wood et al (op cit). They used five-term problems presented textually and found that subjects began by reporting the use of some form of image device but with practice they quickly moved on to elimination strategies. These elimination strategies while providing an effective solution for superlative questions, could not be used to solve comparative questions and Wood et al found that practised subjects, unlike their novice counterparts were quite unable to answer unexpected comparative questions presented at the end of problems.

These results would appear then to offer strong evidence for some form of domain specific expertise factor in the solving of the transitivity problems analagous to that identified in the state-space experiments. Furthermore it would seem that the spatial image device is a rather unsophisticated and ineffective method of solving transitive inference problems and probably not the one which we would want to consider as a particularly expert response to the problem? How true is this?

The Spatial Image and Expertise.

With careful consideration, it becomes clear in fact that the expertise generated in these experiments is not really of the same type as that obtained in the state space experiments. Certainly it is the case that the expertise generated in the transitivity experiments produces dramatic transfer and seems to be limited to problems of the same class. But what has become confused in these studies is precisely which class of problem is being solved by the subjects. It was argued at the outset that the three-term problems could be classified on the basis of the fact that they all required a certain logical inference to be made if the solution was to be attained. This logical inference involved the comparison of two items via a middle term. Under this definition we could also recognise four- and five-term transitivity problems of the type used in the developmental literature to be of essentially the same class. But now consider the expert strategies described by Quinton and Fellows and Wood. Here subjects are no longer required to make such inferences. Because they have access to the question before reading the premises, they only need ever compare two items within any given premise. There is no need for them to combine the premise information in the way entailed in the truly inferential class of problems. Thus it would seem that the paths-to-expertise described here are only made possible if the subject is given the opportunity to implicitly redefine the class of problem which he is solving.

As is clear from the work of McGonigle and Chalmers (1984), there is a lesson to be learnt here for researchers interested in studying inference. Many studies of inference have used procedures which allow simultaneous access to premises and question, and many studies have allowed subjects to practice on "warm-up" problems (Hunter 1957, Clark 1968, Jones 1970, for eg). But these are precisely the conditions that promote the non-inferential strategies identified by Quinton et al. These studies may thus be providing data that has very little to do with transitive inference per se. Lending support to this argument is a study by Potts and Sholz (1975). This revealed that subjects required to solve problems when premises and question were simultaneously available produced a quite different profile to subjects who were required to read the premises before reading the question.

In the developmental literature there have been similar confusions. Bryant and Trabasso (1971) for example, claimed that children as young as four years could make transitive inferences provided certain experimental procedures were used. McGonigle and Chalmers(1977) however, showed that with these procedures even squirrel monkeys could be shown to make apparently transitive choices, but these choices could equally be explained by a model which did not require the coordination of relations. Thus such procedures may only show some form of ´pseudo-transitivity´.

It would seem then that a study of expertise in the solving of inference problems must be careful to ensure that the experimental procedures used to study such expertise do not in fact allow the subject to avoid making inferences altogether. This can be done by using procedures which do not present premises and questions simultaneously or which use comparative rather than superlative questions. There have been experiments using such procedures but while these studies have demonstrated that subjects tended to use some form of spatial image device in these situations, they have not looked at the effects of practice on the use of this device. It is not clear from these studies whether the image device is a relatively sophisticated and expert response to the problem environment or whether subjects quickly abandon this strategy in favour of some other inferential device.

In conclusion then, it would appear that the research on inference makes it clear that the transitivity tasks would be a useful domain in which to study domain specific expertise. There are affluent psychological measures available to the experimenter and already some psychological mechanisms have been postulated to account for performance on these tasks. However there has been no proper study of transfer effects in this task-domain and consequently there has been no consideration of expertise factors in relation to such mechanisms. The experiments reported in the experimental chapter that follows are an attempt to redress this imbalance.

CHAPTER FIVE

SOME EXPERIMENTS WITH TRANSITIVITY TASKS

Experiment Three: Transfer with 3- and 5-Term Series Problems

In the light of the literature reviewed in the previous chapter it is clear that the first aim of this experiment must be to establish clear evidence for the existence of an expertise factor akin to that identified in the state-space experiments. Thus we need to examine the performance of subjects solving transitivity problems in a multiple-problem paradigm which does not promote the use of non-inferential strategies, of the type identifed by Wood (op cit). This can be achieved in one of two ways; Either subjects can be given comparative questions (ie. 'Is D>B?'), as in the studies of Bryant et al (op cit), or alternatively subjects can be given superlative questions (ie. 'Who is tallest?') provided that, these are not presented simultaneously with the premises, but are presented after the premises. In this way, subjects cannot use the question to guide their processing of the premise information. In this study, superlative rather than comparative questions are used for two reasons; Firstly, nearly all of the adult literature on transitive reasonig uses these questions, but none has examined the importance of the relative ordering of question and premises for the development of expertise. Secondly, as part of this study we shall be using 2- dimensional problems which do not lend themselves easily to being associated with comparative questions.

Using this type of procedure, we need to establish whether any transfer effects occur and we need to use protocol analysis to investigate the problem solving strategies used by subjects. In this way we hope to examine whether subjects report using some form of spatial image device as did Huttenlocher´s subjects and to assess the status of such a device with respect to growing expertise; Will the use of such a device be limited to novice performance as in the case of the study of Quinton and Fellows (op cit), or will this new procedure generate a form of expertise which relies on the use of a spatial image device?

To this end the first experiment uses two types of problem. The first type is the usual three-term series problem described in the previous chapter and used by many researchers in the field. The second type of problem is slightly different. This second type consists of a set of five-term problems which use relational terms in two dimensions (for eg A is left of B, B is left of C, D is above C, E is rightof C). This second set of problems was included for the following reasons.

It has been argued strongly (Foos 1980), that subjects solving this type of spatial problem use a spatial image device essentially the same as that reported by Huttenlocher. This being so then, the second set of problems would be of use as a means of testing the transferability of any expertise generated on the three-term tasks. For example, if subjects solving the three-term problems quickly abandon the spatial image device in favour of some more task-specific inferential strategy then they would not be expected to show any improvement over novices on the solving of the five-term problems. On the other hand, if the effects of practice on the three-term subjects was to actually develop the spatial image strategy, then we might expect to find some positive transfer.

Both sets of problem were used in conjunction with superlative questions presented after the premises had been read, and the method of protocol reporting was used to probe for the presence of the image strategy. In particular subjects were required to make a total of 12 protocol reports after varying numbers of problems had been solved. Using this method it was hoped that changes in strategy as a function of growing expertise might be identifiable.

Method

Task Materials

32 inference problems were used in this experiment, 16 three-term series problems of the type used by Huttenlocher, and 16 five term-series problems with relational terms in two dimensions similar to those used by Foos.

The three-term problems consisted of two sets of the 8 regular comparative forms of the problems as in table 5.1.1. The relational terms used were ´taller than´, and ´shorter than´. Relational terms were randomly assigned to the first set of 8 forms with the constraint that 4 of each relation occurred in the first set, and the relational term used in the other set of 8 problems was chosen so that any given problem in this set used the relational term not used by its corresponding version in the first set. The items consisted of Christian names. Of the two sets, 4 used female names and 4 used male names. The names were selected at random from a pool of 16 of each sex, except for the constraint that no two names could ever occur in a single problem together more than once. To make the analysis easier each of the names began with a different initial letter.

		match problems	Non-match problems
A is taller than B	B is taller than C	A is to the left of B	A is to the left of B
B is taller than C	A is taller than B	B is to the left of C	C is to the left of D
		C is to the left of D	B is to the left of C
A is taller than B	C is shorter than B	B is to the left of C	C is to the left of D
C is shorter than B	A is taller than B	C is to the left of D	A is to the left of B
		A is to the left of B	B is to the left of C
B is shorter than A	B is taller than C	B is to the left of C	A is to the left of B
B is taller than C	B is shorter than A	A is to the left of B	C is to the left of D
		C is to the left of D	B is to the left of C
B is shorter than A	C is shorter than B	C is to the left of D	C is to the left of D
C is shorter than B	B is shorter than A	B is to the left of C	A is to the left of B
		A is to the left of B	B is to the left of C

Table 5.1.1 The 8 regular forms of the three- term and 5 term problems (major axis only). Letters correspond to Christian Names (Alan Bob and Carl for eg.)

For each of the two sets of 8 regular forms the premises were combined with one of the two superlative questions: ´Who is tallest?´ or ´Who is shortest?´. Once again for one set of 8 the questions were randomly assigned with the constraint that 4 each of the two forms were presented. For the other 8 the problems were assigned the question not assigned to its corresponding problems in the first 8. Thus there were 16 problems two each of the 8 regular forms differing with respect to question type and, in some cases, the sex of the names used as items. The 16 problems were presented as two blocks of 8 regular forms although for the subject there was no indication of the end of a block.

The composition of the five-term problems is more complex to describe. Again the 16 problems consisted of 2 sets of 8 forms differing according to the question type, and the sex of the names used as items. The 8 forms comprising each set were the 8 basic forms described by Foos and reproduced in table 5.1.1. As can be seen this typology is somewhat different to that of the three-term problems. For the three-term problems the major differences between problem forms is the order of presentation of premise information within each premise. For the five-term problems the major distinction is the order of presentation of problems across each premise. This is congruent with Foos´ own experimental emphasis and his distinction between ´match´ problems in which every successive premise has one term that is mentioned in the immediately prior premise, and ´non-match´ problems in which successive premises may not necessarily have any items in common. This does however lead to severe counterbalancing problems when only 16 problems are used.

Foos used 4 relational terms ´north of´, ´south of´, ´east of´, and ´west of´ and his terms were common words such as ´farm´, ´lake´ etc. From Foos´ report it is not clear how he combined these relational terms with the 8 problem forms. Even for Foos´ own experiment a full counterbalancing is not possible within the number of problems he used. In our case with only 16 problems it is out of the question. But the major problem-type effect recorded by Foos was between ´match and non-match problems rather than to do with the relational terms used or with the orientation of the major axis in the problems. Thus it was decided to use problems whose major axis was always horizontal and whose horizontal relation was the relational term ´left of´. Furthermore the fifth-term in the series was added as either the second or the third premise of the problem and was linked to an end-term and a middle-term of the major axis equally often by the use of either the relational term ´above´ or ´below´.

This is congruent with Foos´ own procedure as far as can be determined. The decision to use the relations ´left of´, ´above´, and ´below´, was taken in order to avoid the difficulty that subjects might have with remembering the cardinal points of the compass, particularly east and west, (Foos´ subjects were given score-sheets with these orientations clearly marked). Also, this would avoid lengthy and cumbersome questions such as ´Who is the southernmost?´. The use of conventional spatial relations is consistent with the earlier work cited by Foos (1980), with the study of Wood et al (1974) and that of Mani and Johnson-Laird (1982) discussed in the next experiment.

As with the three-term problems described above the items used in these problems were Christian names and they were counterbalanced across the problems in the same way as the three-term problems.

Design

Subjects were randomly assigned to 2 groups. Group 1 (n=16) received the 16 three-term problems first followed by the 16 five-term problems. Group 2 (n=14) received the two types of problem in the reverse order. The design is summarised in table 5.1.2.

	first problem set	second problem set
Group 1	1-dimensional 3-term	2-dimensional 5-term
Group 2	2-dimensional 5-term	1-dimensional 3-term

Table 5.1.2 Design of experiment 3

Within each of the 16 problem sets the order of presentation of the problems was fixed. In both cases however the first 8 were instances of the 8 regular forms of table 5.1.1, the second 8 problems solved were a further 8 instances of the same forms presented in the same order.

Subjects

The subjects were 30 3rd-year undergraduates from the department of psychology, University of Edinburgh, and participated in the experiment as part of their practical course and were not paid for their services. Subjects of both sexes were randomly assigned to groups.

Procedure

The problems were presented to subjects on a monitor screen controlled by an APPLE II microcomputer. After preliminary information about the experiment (which did not include warm-up trials or examples of the type of problems to be solved) subjects were required to press the <space> bar of the computer keyboard and the premises of the first problem were presented on the screen. All of the premises of the problem were presented simulataneously for a fixed duration of 12 seconds for the three-term problems and 24 seconds for the five term problems (6 seconds per premise). This latency was controlled by the computer. After this time the premise information was replaced by the appropriate question. The subject was then required to press a key while at the same time verbalising the answer to the question. The key-press stopped a timing routine controlled by the computer which timed the duration between the onset of the question and the key-press. The subjects' verbal responses were recorded by the experimenter who was in the same room as the subject throughout the session.

Each successive problem was presented by the computer in response to the subject pressing the <space> bar and the subject was given no indication of how many problems he would be required to solve or any warning of the change in problem type after the first 16 problems. Subjects were informed that their question answering time was being recorded and that they were to answer the problems as quickly as they could without making unnecessary errors.

The problems were presented to subjects on a monitor screen controlled by an APPLE II microcomputer. After preliminary information about the experiment (which did not include warm-up trials or examples of the type of problems to be solved) subjects were required to press the <space> bar of the computer keyboard and the premises of the first problem were presented on the screen. All the premises of the problem were presented simulataneously for a fixed duration of 12 seconds for the three-term problems and 24 seconds for the five term-problems (6 seconds per premise). This latency was controlled by the computer. After this time the premise information was replaced by the appropriate question. The subject was then required to press a key while at the same time verbalisisng the answer to the question. The key press stopped a timing routine controlled by the computer which timed the duration between the onset of the question and the key-press. The subjects verbal response was recorded by the experimenter who was in the same room as the subject throughout the session.

At intervals throughout the session the experimenter interrupted the subject to ask him to provide a written protocol report describing how he was solving the problems. Every subject was required to give 6 such protocol reports for each set of problems giving a total of 12 reports for each subject. These were taken after the 1st, 4th, 8th, 9th, 12th, and 16th problems of each set. After each of these particular problems had been solved, a buzzer sounded and the words "take a break" were presented on the screen. At this point the experimenter asked the subject to write down a few words about how he was solving the problems. When the subject had done this, he pressed the <space> bar twice, to receive the premises of the next problem. Although subjects were informed at the beginning of the session that they would be asked to provide protocol reports they were not informed as to when this would occur.

Results

Table 5.1.3 presents the mean question answering time for both groups of subjects solving both types of problem.

	first problem set	second problem set
Group 1	85.76	121.52
Group 2	189.07	84.11

Table 5.1.3 Group comparisons of mean question answering time for both sets of problems

An analysis of variance (BMDP P2V) on these data revealed a significant difference between the two groups ($F_{1,28} = 4.56$; $p = 0.0417$), a significant effect of problem set ($F_{1,28} = 18.98$; $p = 0.0002$) and a significant group by problem set interaction($F_{1,28} = 99.2$,; $p < 0.0001$). T-tests for group effects on problem sets solved first and second revealed that there was a significant effect of group for the problems solved as the first set ($t = 2.1x10^{-6}$; $p < 0.01$), but no differences between the groups on the problems solved as the second set ($t = 1.3x10^{-3}$; $p > 0.05$). An inspection of table 5.1.3 reveals that the five-term series problems solved as the first set of problems by group 1 produce a much slower solution time than the three-term set solved first by group 2.

As can be seen however this difference is much reduced when the two sets of problems are solved as the second set suggesting that the experience that the group 1 subjects had with the three-term problems solved as their first set allowed them to solve the five-term problems more easily than group 2 solving them as naive subjects. In support of this, t-tests reveal that there is a significant difference in solution time between the five-term problems solved in position 1 and position 2 ($t = 1.1 \times 10^{-3}$; $p = 0.0012$) but no significant difference between the three-term problems solved first and second ($t = 6.8 \times 10^{-1}$; $p = 0.68$). These results suggest a positive transfer between three- and five-term problems but only when the five-term problems are solved second.

Turning now to an analysis of the effects of practice within each set of problems, figure 5.1.1 (overleaf) presents the mean question answering time as a function of the number of problems solved for both sets of problems and both groups. The 16 problems of each set have been collapsed into blocks of 4 for convenience of presentation but analyses of variance treated all 32 problems as separate. In addition to the group and set effects reported above, these analyses revealed a significant effect of practice ($F_{15,420} = 7.59$; $p < 0.0001$) and a significant three-way interaction between practice, set and group ($F_{15,420} = 2.07$; $p = 0.0105$).

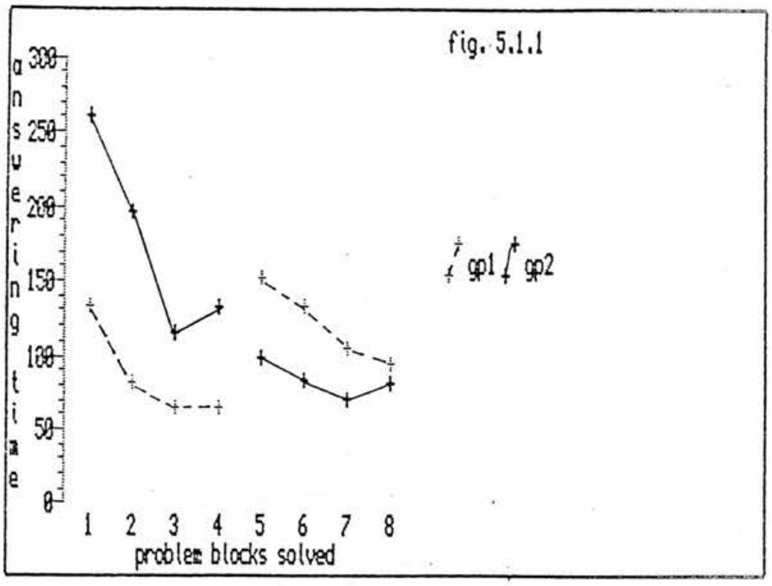


Figure 5.1.1. Group comparisons of question answering time (50ths of second), as a function of number of problems solved. Problems are presented as 8 blocks of 4. Block 5 marks the start of a new set of problems for both groups.

Separate analyses on each problem set for each group revealed a significant effect of practice for the five-term problems solved as the first set ($F_{15,195} = 3.48$; $p = < 0.0001$), for the three-term problems solved as the first set ($F_{15,225} = 2.58$; $p = 0.0014$), for the five-term problems solved as the second set ($F_{15,225} = 2.28$; $P = 0.0052$), but not for the three-term problems solved as the second set ($F_{15,195} = 1.6$; $P = 0.077$).

These results, in consideration of figure 5.1.1, are interpreted as suggesting there was significant improvement in solution time as a function of practice within each set of problems. The exception- the three-term problems solved as the second set- may be attributable to floor effects. Turning now to a consideration of the proportion of problems answered correctly; It was found that the proportion of errors is, by and large low. The average error rate for the three-term problems is about 10% and for the five-term problems, about 30%. There is some indication of the frequency of errors dropping as a function of practice in a way that is analogous to the reduction in solution time presented in figure 5.1.1. But because of the low frequency of errors, no statistical analysis was carried out on these data. It now remains to be seen whether the protocol reports reveal anything about the psychological changes that underwrite these transfer effects.

After consideration of the literature in general and the particular emphasis of this experiment, the protocols were examined for 5 major categories of report. Appendix D contains a detailed description of these categories, along with an explanation of the categorisation process supported by examination of selected protocols.

First of all, the protocols were examined for evidence that the subject used some form of mental image. These reports fall into two categories; GLOBAL IMAGE and DIMENSIONALISED IMAGE. Global image refers to fairly gross reports such as "I made a mental picture in my minds eye". These reports do not provide enough information to decide whether the subject is literally picturing the premises as they are written, whether he has some very concrete image of a set of people (Egan and Grimes-Farrow 1982), or whether he is using the kind spatial image device described by Huttenlocher (op cit) which is characterised by a spatial array of the names along a single dimension corresponding to the relative heights of the people. In contrast, dimensionalised image refers to reports which are clearly instances of a Huttenlocher-type of image such as, "I made a mental picture with the tallest at the right, smallest at the left and so on."

Mynatt and Smith (1977) make a distinction between the use of image strategies and the use of VERBAL REHEARSAL strategies, and the 3rd category for which the protocols were examined was the use of rehearsal strategies. Reports such as, "I made a verbal list of the names and repeated them over to myself, tallest first" would qualify for this category. Many subjects quite simply reported using "verbal rehearsal of names".

The 4th category- SCANNING- refers to reports indicating that the subject is not reading the premises in the order which they are presented but instead is searching the premises for certain kinds of information. In particular searching for premises which contain information that can be added directly to the existing representation. Thus if a subject is presented with premises such as ´A>B, C>D B>C´, as he might be as part of a five-term problem, he might report for example, "take first name, attach second to it, scan for any other occurrence of first name etc. ie. taking one name and attaching others to it rather than having say, two pairs of names and later joining them up".

The final category is ELIMINATION. This is to be distinguished from the kind of elimination discussed by Wood (op cit, see introduction). In these reports the subject does not search the premises for the answer to the question- he could not because he doesn´t receive the question until he has read the premises. Instead he reports using a spatial image but remembering only the end-anchors of the array. He can do this because although he does not know what question he will be asked, he knows that it will be a superlative question such as ´Who is leftmost?´ or ´who is tallest?´ and thus he need only remember the extreme items, in the words of one subject; "..only remembering the outside names so less names to remember. This is because I know the limited number of questions being used....hope the questions don´t change!"

Each of the 12 protocols from each subject was inspected for these five categories of report. The categories were not mutually exclusive and a single report can contribute a data point to several categories, as in the case of; "I made a visual image of tallest on the left, smallest on the right (category 2), then rehearsed the names (category 3)". These 5 categories accounted for over 80% of all reports. Other types of report were the use of mnemonics such as remembering the initial letters of the names mentioned in the premises, and a small number of reports of remembering the premises verbatim (mainly in the three-term set), and also of non-directive reports such as "I made a mistake because I was feeling tired".

Figure 5.1.2. (a-d) on the facing page, presents the frequency of occurrence of each of the five categories of report. This is broken down into problem set (3- and 5- term) for both of the groups. Thus, figure 5.1.2a gives the data for group 1 solving their first set of problems- the three-term problems, and figure 5.1.2b the same group solving the five-term problems which they solved subsequently. Similarly for group 2 in figures 5.1.2c and 5.1.2d. Note also, that for each group solving each set of problems, the data is broken down as a function of practice. Thus in figure 5.1.2a, the distribution of protocols over the categories is presented in 6 blocks. The first block corresponds to the protocol reports which were recorded immediately after the subjects had solved their 1st problem, the second block corresponds to those reports taken immediately after they had solved their 4th problem of that set and so on, until we come to the last block of figure 5.1.2a which corresponds to those reports taken immediately after the subject had solved all 16 of the three-term problems. Similarly for figures 5.1.2 b,c, & d.

The first thing to note is that in all groups for both sets of problems, the frequency of reports of some form of image strategy was high and appeared very early in the experimental session. In fact 75% of subjects reported using such a strategy even on the protocol report taken immediately after their first problem. Furthermore, the frequency of such reports remained fairly constant over subsequent reports, on their last report 80% of subjects reported some form of image device. It seems then that the use of the image device occurs in response to the earliest reports and is not given up in favour of some other strategy. Instead, other categories of report occur in addition to the stable image reports and as an elaboration of them. The subjects do not appear for example, to begin by using some form of image device and then abandon this in favour of a purely verbal rehearsal strategy. Instead the subject begins by reporting the use of an image and subsequently he reports using an image plus rehearsal as in: " I ordered the people in a line, tallest at the left", and subsequently: "picture the names in a line and verbally repeat the names".

The dimension of this elaboration differs as a function of problem type and group. For the three-term problems solved first (group 1: fig 5.1.2a), subjects appear to begin by reporting the use of a global undifferentiated image strategy which is quickly refined into a dimensionalised image. Thus several subjects in this group reported in their early protocols; "I made a mental picture of the names" and later reported; "I ordered the names from tallest to shortest, tallest on the left, shortest on the right". In addition, subjects augment this by the use of rehearsal as described above. There is very little evidence for the use of scanning and elimination for these problems. For the three-term problems solved second (group2: fig 5.1.2d), the profile is similar if somewhat attenuated.

For the five-term problems solved first (group 2: fig 5.1.2.c) there was an addition of rehearsal to the basic image as with the three-term problems. But for these problems there was a large number of reports indicating that this was even further elaborated by the addition of scanning and/or elimination. Thus one subject reports the following on his first report; "Picture of relationships" (global image), and the following on his penultimate report; "...slot people into a position from left to right (dimensionalised image), then drop the middle names (elimination)". For the five-term problems solved second, again there is a high incidence of some form of image device elaborated by rehearsal and scanning, but there is less evidence of the use of elimination. It is noteworthy that protocols from these five-term problems generally do not show evidence of the dimensionalised image that features so strongly in the three-term problems, rather image reports tended to be of a more global nature. This is presumably due to the fact that the relational terms used in the premises of these problems were explicitly spatial- it would be somewhat pedantic for the subject to report, "placing the leftmost item at the left" and so on.

Considering this strategic elaboration in purely quantitative terms, it is possible to examine the extent of this elaboration and whether there is more elaboration shown when subjects are required to solve the more difficult five-term problems. This is achieved by counting how many categories each report falls into and comparing this with previous reports from the same subject. If a subject reports using a dimensionalised image in his first report and then reports using a dimensionalised image plus rehearsal in his second report, and he reports no change subsequently, then he is scored as having made 1 strategic elaboration during his solving of the problems and so on. Figure 5.1.3 presents the percentage of each group of subjects who showed varying amounts of strategic elaboration. This is presented as a function of problem set.

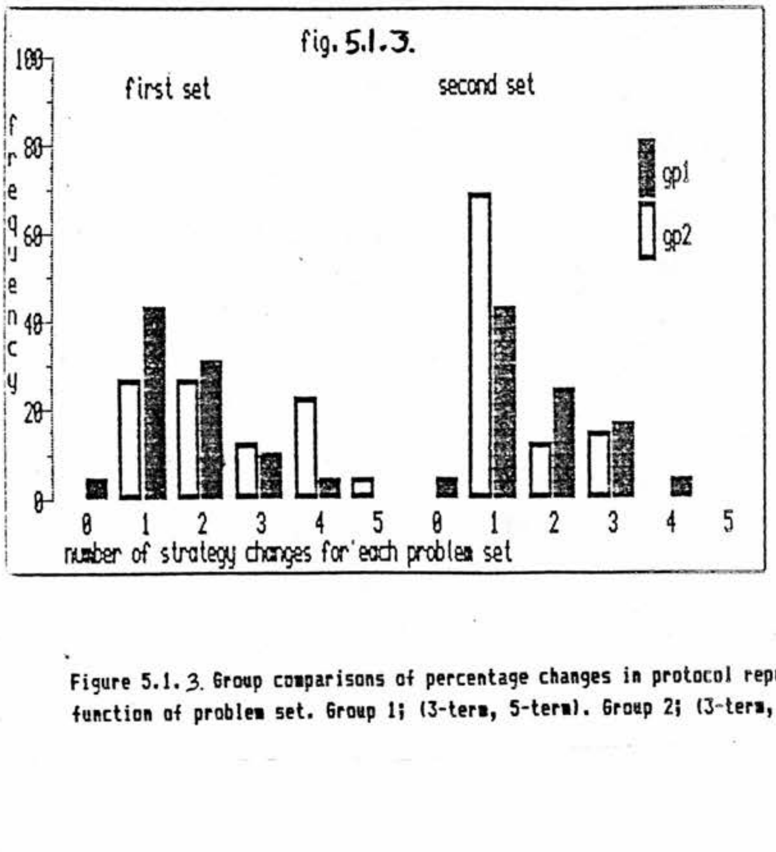


Figure 5.1.3. Group comparisons of percentage changes in protocol reports as a function of problem set. Group 1; (3-term, 5-term). Group 2; (3-term, 5-term).

As can be seen, in terms of the modal frequencies, most subjects solving the 3-term problems showed 1 or 2 strategic elaborations, whereas most subjects solving the 5-term problems showed between 2 and 3 elaborations with some subjects showing as many as 4 and even

5 elaborations. It seems that the problems which are the most difficult to solve- as determined by solution time (cf. fig. 5.1.1)- are also those for which subjects tend to show the most strategic elaboration.

Finally, it is possible to examine regularities in the sequential course of these strategic elaborations. This is achieved by examining the frequency of occurrence of various patterns of elaboration. Since there are five categories, there are 5!=60 ways of combining all five categories. But in fact, there was a fairly regular syntax to the elaborations. For the three-term problems, most subjects showed the following pattern; global image -> dimensionalised image -> rehearsal. A few subject showed; rehearsal -> global image -> dimensionalised image. For the five-term problems the most common syntax was global image -> rehearsal -> scanning -> elimination, although few subjects reached elimination. Once again, a few subjects showed rehearsal before global image but no subject showed scanning or elimination before global image or rehearsal. Examples of protocol reports obtained in this study are presented in appendix D, along with further details of the classificatory procedures.

Discussion

The solution time data leave no doubt that there is an expertise factor involved in the solution of these inference problems, and the protocol reports have proved to be a useful source of information concerning the psychological mechanisms involved in such expertise.

These reveal that nearly all subjects use some form of spatial image device in order to solve these problems. The use of such a device appears early in the protocol reports. Even after their first problem most subjects report using some form of image. Furthermore the dimension of strategic change associated with reduction in solution times, unlike previous studies (Wood et al op cit) is not a shift away from the use of this device. Subjects persisted in the use of the spatial image throughout the experimental seesion. The dimension of strategic change appeared rather to be an elaboration of the basic image device.

One elaborative dimension was the addition of strategies dedicated to reducing the memorial strain of the task. Thus subjects report using mnemonics such as rehearsal, and remembering only the end-anchor items of the premises. The fact that such reports were more frequent for the five-term problems is indicative of the greater strain that these problems place on memory.

Such supportive mnemonics are not however to be regarded as a domain specific aspect of expertise. Indeed the use of such mnemonics can be seen as a much more general problem solving skill involving subjects´ ´metacognitive knowledge´ of their own memory processes. A much more likely candidate for a domain specific component is the use of scanning. These reports suggest that subjects do not inevitably process premise information in a bottom-up way in the order in which they are presented on the screen, rather they actively search for items that can most easily be integrated into the spatial image. They report for example searching for items that can be linked directly onto an item that is already established in the image.

In support of this Foos (op cit) has argued that a major contribution to problem difficulty in his experiments was the need to hold in memory, items that could not be linked immediately onto the array under construction. So for example, the problem; "A>B, B>C, C>D, D>E" is easy to solve in comparison to the problem; "A>B, D>E, C>D, B>C" in which the pair; D,E cannot be integrated into the representation until all the premises have been read. It would seem that scanning allows subjects to overcome this difficulty.

This particular version of scanning would not be appropriate for the three-term problems, since because they have only two premises there is always a common term in the new premise. However, the fact that subjects report building the image in a preferred direction, namely starting with the tallest, would suggest that subjects may scan the premises for items that are congruent with this preferred direction of working. This interpretation of scanning, in accordance with McGonigle and Chalmers (1984), forces a distinction between the representational space used to solve these problems and the information processing strategies dedicated to translating problems into that space.

The fact that subjects report using an image representation to solve both types of problem presented in this experiment suggests that this device is a quite general background resource, capable of being used in a number of different problem solving situations (see for example, Lakoff and Johnson (1980) for evidence of the use of spatial metaphors in other situations). The information processing strategies such as scanning on the other hand, dedicated as they are to translating between individual problems and this general representational space, must be domain specific. The very fact that the three-term problems used non-spatial relational terms while the five-term problems used explicitly spatial relations, yet at the same time, subjects report using a similar representational space for both problems, is evidence for the specificity of these translation strategies.

Taking this as a working hypothesis about the nature of the expertise involved in the solving of these problems what further work needs to be done? Firstly it will be necessary to come to a clear understanding of the characteristics of the representational resource involved here. If it is a quite general resource how does a subject come to recognise when a particular problem domain is solvable by use of this resource? What aspects of the problem domain constrain the generality of this resource? We have already observed that the relational information does not have to be inherently spatial. It is also questionable whether or not this information even has to be fully ordered. In his experiment, Foos found that subjects given indeterminate premise information such as "A is left of B, C is below B,D is right of C", would interpret this information as if D was directly below A.

While this result would not be surprising given the suggestion of McGonigle and Chalmers (op cit) that the image is imposed upon premise information as an interpretive schema and as an ´effort after meaning´, it does conflict with the research of Mani and Johnson-Laird (1984), who found that subjects could not construct a spatial image representation from such indeterminate premises. This is the first issue taken up in the final experiment of the thesis.

As a second consideration in the final experiment it will be necessary to make a detailed examination of the information processing strategies used by subjects to overcome the surface structure variation inherent in this problem domain. In accordance with most other work in this area experiment 3 concentrated mainly on question answering time. It is clear however, given the above hypotheses, that the effects of problem variation should be reflected not in question answering time but rather in premise reading times.

In particular if one were to establish that the image device is indeed asymmetrical with respect to preferred direction of working, it is possible to predict which surface forms of a transitive inference problem will require conversion in order to make them congruent with this preferred direction of working. This extra processsing should be observed as greater premise integration times for these non-preferred forms. These ´figural effects´ as they have been called (Johnson-Laird 1984), are common in the research literature. Indeed as we have already seen, they provide the mainstay of evidence for Huttenlocher´s model of reasoning (see also Clark and Hunter op cit).

What is not known however, is what happens to these figural effects as the subjects become more expert. On the one hand there is reason to believe that figural effects should subside with practice. After all, the operational definition of domain specific expertise put forward in this thesis can be stated as; that which renders formally equivalent problems psychologically equivalent. On the other hand if it is indeed the case that even expert subjects used a dimensionalised spatial image strategy surely the adverse effects of incongruence must persist? This is the second issue taken up in the final experiment.

Experiment Four: Representation, Strategy and Expertise

The ´top-down´ view of problem solving developed thus far suggests that background representational resources such as the spatial image device are used as an interpretive schema for incoming information. As McGonigle and Chalmers (1984b) point out this view has good analogies in the field of perception in which redundant information in ´good figures´ such as the circle allows the perceiver to infer the presence of a whole circle even when it is partially occluded.

This view of the role of the image predicts that even when the premise information is incomplete or ´indeterminate´ the subject should be able to use the image device to impose an ordering structure on that information. But recent work by Mani and Johnson-Laird (1982) has suggested that this might not be the case. In their study they found that subject readily used some form of spatial imagery to interpret ´determinate´ problem information such as;

> The knife is to the left of the plate
> The plate is to the left of the fork
> the plate is below the spoon

But for indeterminate versions of the same problem in which there is some ambiguity as to the appropriate spatial layout of the items, such as;

The knife is to the left of the plate

The plate is to the right of the fork

The plate is below the spoon

they found that subjects did not seem to use a spatial image representation. Rather they retained the premise information in a propositional form close to the surface structure format of the premises.

One of the problems with this study though, is that there may have been some ambiguity on the part of the subject, as to what the nature of the problem solving tasks actually was. On the one hand subjects in this study were required to verify whether a diagram was consistent with the premise information and on the other hand they were given incidental recall tests. The results of the state space experiments made it clear, that the way in which the subject interprets the <u>functional goal</u> of the problem solving experiment has a major effect on what kind of strategy he adopts. The use of both memory and drawing tasks in Mani and Johnson-Laird's experiment may mean that the subject does not have a clear idea of what is required of him in this experiment.

In the experiment to be reported here, both determinate and indeterminate problems are used and the functional goals of the problem solving task are directly manipulated. This is achieved by varying the type of question that accompanies the premises. In one condition both determinate and indeterminate premises are presented in conjunction with a set of comparative questions of the type used in developmental research by Trabasso et al (op cit). In this condition subjects are required to make comparisons between all possible pairings of items in a five-term series (ie. Is A>E?, Is B>C? etc.). It is argued that this question set requiring as it does that the subject has a ´full perspective´ on the premise information available, will encourage subjects to use a spatial image device as an interpretive schema for the premise information in both determinate and indeterminate cases. This question set also allows for the measurement of inference distance effects which Trabasso claims are indicative of the use of spatial image devices (cf chapter four), hence it will provide evidence for the use of such devices additional to protocol reports.

In a second condition subjects are required to solve indeterminate problems, but this time not with a full set of comparative questions, but with the more usual set of two superlative questions (Who is tallest/shortest?). Furthermore these are presented before the premise information. This kind of procedure, it will be recalled, is precisely that which encourages, not the use of a spatial image,but rather some form of elimination strategy of the type identified by Wood. This question set, requiring as it does, only a ´partial perspective´ on the premise information should be unaffected by the determinacy of the premises. This condition will also allow for comparison between the amount of reduction in solution time produced by the use of the two kinds of strategy.

It is hoped that this manipulation of the functional goals of the problem solving task will encourage a clear distinction in the protocols between those subjects using a spatial image device and those using some form of elimination strategy. If there is a high percentage of subjects using spatial images with indeterminate problems this will be good evidence for the view of the image as a quite general representational device. It is then necessary to consider what kinds of domain specific information processing strategies subjects need to deploy in order to make the various surface forms of the premises congruent with such a general schema.

With respect to this issue, it will be recalled that Huttenlocher´s model makes certain predictions about the relative difficulty of various forms of the three-term problems in accordance with the number of operations that are required to translate the information into the spatial image. So for example the premises; "A is taller than B, C is shorter than B" were said to be easier than problems of the form "B is shorter than A, B is taller than C", since in the latter form the first premise is incongruent with preferred direction of working and the new term of the second premise is introduced as the grammatical object.

More recently Potts and Sholz (1975), and also Foos (op cit) have provided evidence that Huttenlocher´s predictions require some modification. In particular they suggest that it is not the grammatical status of an item that is significant but rather the congruence of the relational term describing a new item with the position of that item in the array. Thus they suggest that a new item "school" introduced in the premise, "The lake is north of the school", will be more difficult to process if in fact the school is the southernmost item in the array. This modification is of particular interest since like the preferred direction of working hypothesis, it emphasises that it is the status of the item with respect to the subjects representational space that is important rather than some absolute linguistic characteristic of the premise per se.

Along with the direction of working hypothesis the congruence hypothesis allows us to rank different problem forms with respect to their expected difficulty, so for example in the problem;

D is taller than E

C is taller than D

B is taller than C

A is taller than B

the premise information is presented in a way that is inconsistent with most subjects preferred direction of working which is tall to short. It should thus be a relatively difficult problem to solve. The following problem;

E is smaller than D

D is smaller than C

C is smaller than B

B is smaller than A

should be even more difficult since not only is it inconsistent with subjects preferred direction of working but also the relational terms used to introduce new items are incongruent with that item´s placement in the actual array.

Such difficult problems can be seen as requiring extra processing operations in order to convert the premise information into a form more easily assimilated into the representational space. Such processing operations provide the bridge between the specific characteristics of a particular problem that is to be solved and the architecture of the representational space that the subject employs to solve the whole class of related problems.

In this experiment, these figural effects are examined by the use of a procedure which allows the subject to control his own premise reading time. In this way it is possible to examine the reading time profiles for various premise figures. If indeed it is the case that certain figures require additional processing to make them congruent with the underlying image device, then this would be expected to show up in such measures.

These figural effects should be limited to subjects solving problems by the use of a spatial image, and whether or not such effects remain as expertise grows will be relevant to our interpretation of the nature of domain specific expertise. On the one hand since the image has been shown to remain stable with respect to practice, it would be expected that figural effects must also remain constant irrespective of a general decrease in solution time. On the other hand, if domain specific expertise is to be regarded as rendering formally equivalent problems psychologically equivalent, it would be expected that the general reduction in solution time concomitant with practice should in fact be brought about by a collapsing of such figure effects.

Method

Task Materials

The problems solved by subjects were five-term series problems using the relations ´taller than/shorter than´. Two basic sets were used; a set of 32 determinate problems and a set of 32 indeterminate problems. The determinate set consisted of 8 each of the 4 figures listed below;

i) A>B, B>C, C>D, D>E

ii) B<A, C<B, D<C, E<D

iii) D>E, C>D, B>C, A>B

iv) E<D, D<C, C<B, A<B

In each case the relational term was either 'taller than' or 'shorter than' and the letters were replaced by Christian names as in the previous experiment. It will be noted that the first two problems are consistent with a tall-to-short preferred direction of working, but that problem two introduces items in a way that is incongruent with their final placement in the array. Similarly the third and fourth problems are inconsistent with a short-to-tall preferred direction of working, and problem three introduces problems in a way that is incongruent with their final placement.

The indeterminate set of problems were produced from the determinate set in a way outlined by Mani and Johnson-Laird, In each case the indeterminacy is introduced into the second premise by replacing the middle term brought over from premise 1 by the end anchor of premise 1, thus;

i) A>B, A>C, C>D, D>E

ii) B<A, C<A, D<C, E<D

iii) D>E, C>E, B>C, A>B

iv) E<D, E<C, C<B, B<A

these two sets of problems were combined with two types of question set to produce the two functional perspectives. The partial perspective set consisted of only two forms;

i) Who is tallest?

ii) Who is shortest?

and one of these questions was presented <u>before</u> the subject read the relevant premises, thus allowing for the use of Wood-type elimination strategies based on selective search for the premise information pertinent to the question.

The full perspective set consisted of 7 basic types of question each presented with both ´taller than´ and ´shorter than´ relational terms, giving a total of 14 different forms:

i) Who is taller/shorter A or B?

ii) Who is taller/shorter D or E?

iii) Who is taller/shorter B or C?

iv) Who is taller/shorter B or D?

v) Who is taller/shorter A or D?

vi) Who is taller/shorter B or E?

vii) Who is taller/shorter A or E?

As will be noted these questions require the subject to make comparisons of two items separated by 0,1,2, or 3 inferential steps. In this way it was intended that the full perspective set would encourage the use of a spatial image device and furthermore the use of such a device could be evidenced by the appearance of inference distance effects as suggested by Trabasso. To make up the set of 16 questions a second question 7 was added. The subject was required to answer two of the above questions <u>after</u> reading each of the 32 problems. For each problem one of the questions involved a small inference distance (0 or 1 step) while the other involved a large inference distance (2 or 3 steps). The order of presentation of all questions was counterbalanced across problems.

Subjects

40 Subjects of both sexes took part in this experiment. Subjects were attending third-year practical classes in the department of psychology at the University of Edinburgh. Subjects were randomly assigned to 3 groups.

Apparatus

An Apple II microcomputer was used to present questions and record response times as in experiment 3. In this experiment however subjects were required to press the <space> bar to obtain each premise one at a time.

Design

Subjects were divided into 3 groups. One group solved determinate problems with the full perspective questions, a second group solved indeterminate problems with the full perspective questions and the third group solved indeterminate problems with the partial perspective questions. This is summarised in table 5.2.1 overleaf.

	Problem Type	Question set
Gp 1 (n=15)	Determinate	Full Perspective
Gp 2 (n=12)	indeterminate	Full Perspective
Gp 3 (n=13)	Indeterminate	Partial Perspective

Table 5.2.1 Design of Experiment 4.

Within each group subjects solved the 32 problems in the same fixed order, namely, two figure 1 problems, two figure 2 problems, two figure 3 problems, two figure 4 problems; and this block of 8 was repeated four times. Groups 1 and 2, solving the full perspective questions were further subdivided into 4 subgroups so that the order of presentation of the questions could be counterbalanced. Subgroups 1 and 3 received the questions in the order 1,5,2,6,3,7,4,8 repeated 8 times to give the 64 questions (2 per problem). Subgroup 3 differed from subgroup 1 in as much as the order of presentation of questions within each of the 32 problems was reversed. Subgroups 2 and 4 solved the problems in the order 8,4,7,3,6,2,5,1, but were otherwise identical to the other two subgroups.

This counterbalancing meant that across any one group all 16 question types were presented within the first 4 problems, and thus it will be possible to examine inference distance effects over practice. The relational terms used for the first 32 questions was randomly chosen with the constraint that both questions on any given problem must use the same term. The second 32 questions used the relational term that had not been used on the corresponding question of the first 32.

Procedure

As in experiment 3 subjects were tested individually with an experimenter present to record the subjects verbalised answers to questions. Each subject was informed of the general nature of the experiment but was not given warm-up problems. Subjects were informed that their performance was being timed and were encouraged to solve the problems as fast as was consistent with minimal errors. To this end subjects in this experiment were informed by the experimenter when they had made a mistake in answering the question (in the case of indeterminate problems this meant providing an answer that was inconsistent with any possible interpretation of the premises).

In order to encourage the development of elimination strategies in the partial question group, subjects were informed that they were allowed to answer the question as soon as they were able so to do, even if they had not finished reading all of the premises. (In fact the choice of problem figures was such that subjects could actually answer the partial questions after reading only the first two premises of the problem). In this condition the experimenter also recorded the premise upon which the subject answered the question.

As with experiment 3 it was intended to take protocol reports during the session to provide further information about subjects problem solving methods. It was intended to take 12 protocol reports at varying intervals through the session. Unfortunately owing to time constraints this did not prove possible in all cases. While some subjects provided all twelve protocols the majority only provided 6, at intervals throughout the 32 problems, and some only provided one after the first problem and one after the last. Nevertheless, subjects were encouraged to give a full description of their problem solving methods and whenever a subject reported using a spatial image device then the experimenter asked the subject to specify in which direction the subject constructed the array, i.e. whether from tall to short on a vertical axis with tall at the top or on a horizontal axis with tallest at the left and so on.

As can be seen from the design section, order of presentation of the four premise figures and question form was organised in such a way that each problem occurred in conjunction with each question form at regular (8-problem) intervals throughout the session. Thus by grouping the 32 problems into 4 blocks of 8 it is possible to examine the effects of practice on premise figure and question form (particularly inference distance effects for groups 1 and 2) for each group of subjects.

Results

Three categories of results are reported here. The protocol reports are examined for evidence of both image formation and elimination strategies, and then the question answering time is examined for support for these reports. Finally premise integration time is examined with respect to figure effects and the effects of practice.

Protocol Reports

Table 5.2.2 presents the percentage of reports falling into the two main categories of elimination and image formation.

	Group 1	Group 2	Group 3
Image Formation	87%	94%	38%
Elimination	0%	0%	61%
Elaboration:			
Rehearsal	80%	75%	0%
Mnemonics	53%	33%	23%

Table 5.2.2 Percentage of protocol reports falling into the three main categories

It can be seen from table 5.2.2 that, as predicted, the majority of subjects, in groups 1 and 2 use an image formation strategy which is elaborated by the use of mnemonics and rehearsal as in experiment 3. Mnemonics refers to the use of memory saving devices such as remembering only the initial letters of the items mentioned in the premises. Subjects in group 2 solving indeterminate problems, if anything, show a higher frequency of such reports than those in group 1. In the partial perspective condition however, where subject were again solving indeterminate problems the frequency of image reports was much less and the majority of subjects reported using an elimination strategy. As can be seen the frequency of image reports was however still reasonably high. Of those image reports, 80% occurred in the early reports and were later replaced by elimination reports, the remainder occurred early and persisted throughout the session. A detailed analysis of these shifts is not possible due to the low density of protocol reports provided by some subjects. Of the subjects reporting the use of a spatial image strategy, 98% report constructing an array from tallest to shortest along a horizontal axis, with the tallest at the left.

Question Answering Time

An analysis of question answering time is not meaningful for group 3 since this group received questions before premises, but for the other two groups question answering time can be used to examine inference distance effects. Figure 5.2.1 (overleaf) presents question answering time averaged over the 4 inference distances.

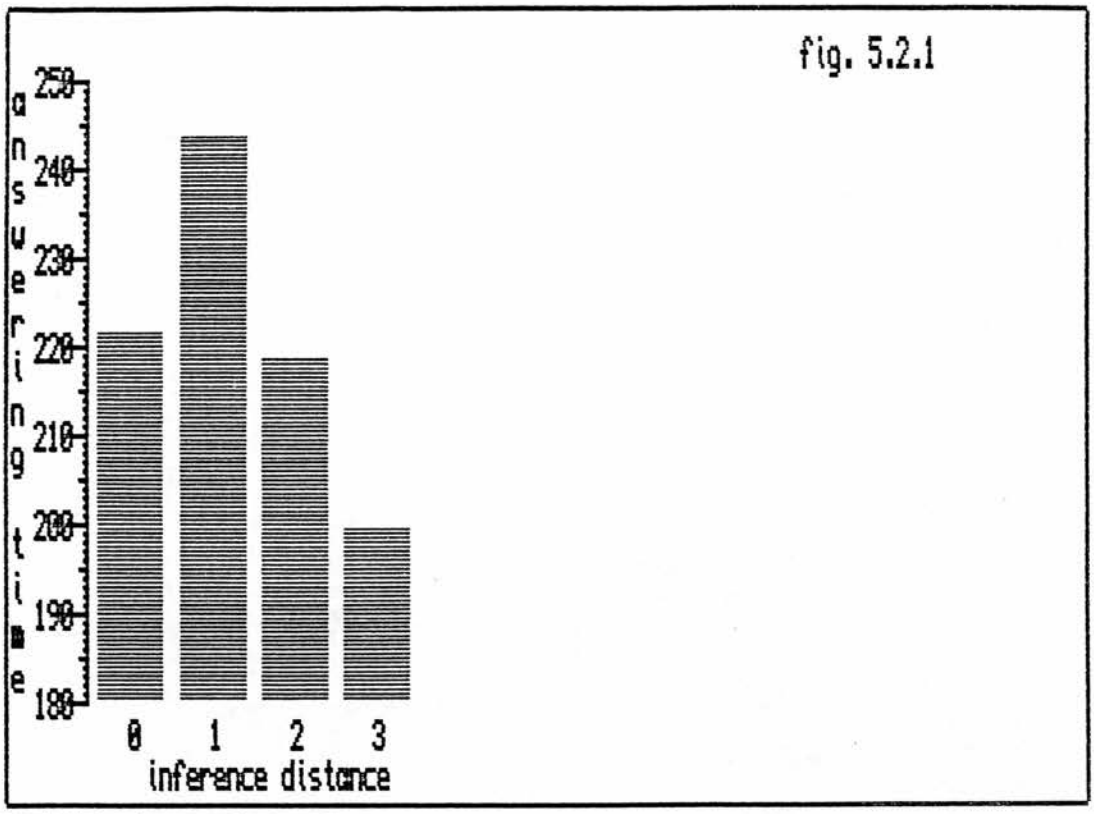


Figure 5.2.1 Inference distance effects (means for groups 1 and 2 only)

A distance of 0 indicates that the two items compared in the question were actually adjacent items and thus were presented in the same premise. A distance of 1 indicates that the two items compared are separated by one other item, a distance of 2, that they are separated by two other items, and 3 that they are separated by three items and are thus the two end-terms of the five-term series.

It will be recalled, that Trabasso argued that the subjects´ use of a spatial image representation could be inferred from the counter-intuitive result that subjects respond faster to items that are separated by a large inference distance than to items that are actually presented together in the premises. It might be expected then, on the basis of the protocol results above, that subjects in groups 1 and 2 would show such a question answering time profile.

As can be seen from figure 5.2.1, this is indeed the case. The faster response time to 0 distance questions than to 1 distance questions can be accounted for by the fact that the 0 distance questions contain a higher proportion of end-anchored questions which tend to be answered faster independantly of inference distance effects. When the inference distance effect is examined as a function of practice as in figure 5.2.2 below, it can be seen that the effect emerges early in practice and remains relatively stable throughout the session, a finding which concords with the protocol reports for these subjects.

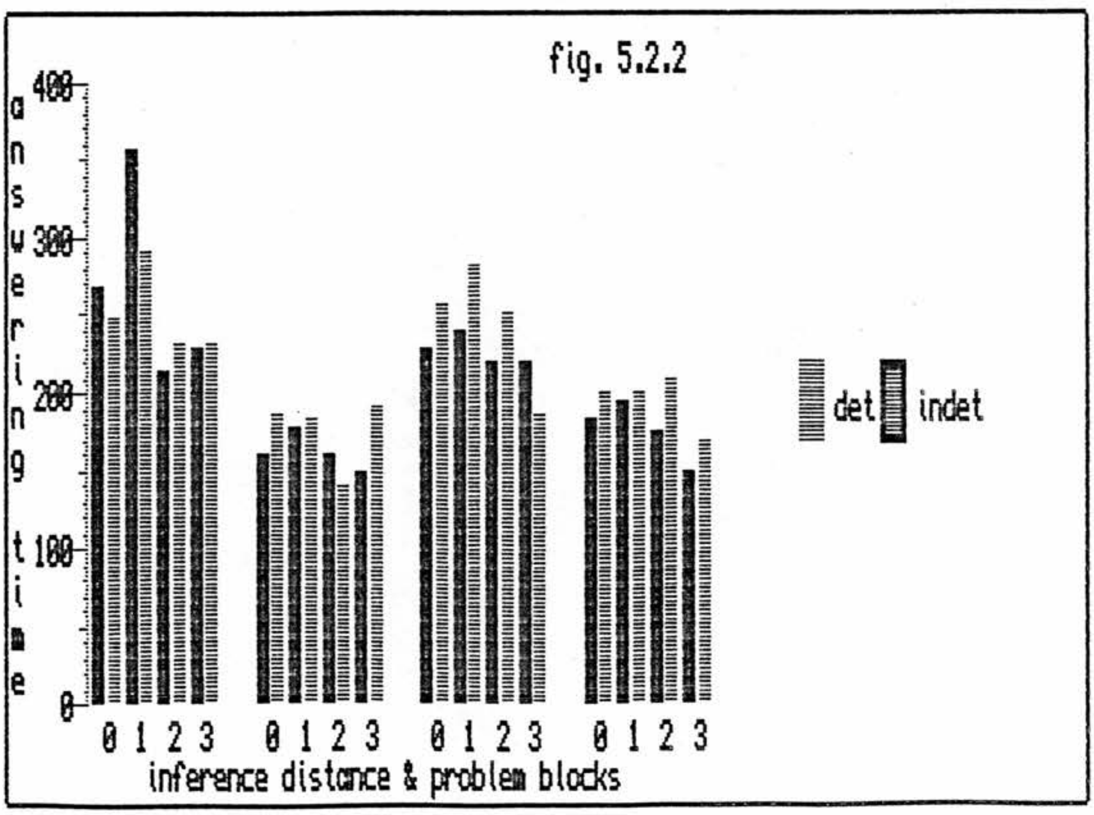


Figure 5.2.2 Inference distance effects as a function of problem blocks solved

An analysis of variance on these data reveals, a significant inference distance effect ($F_{3,69} = 12.29$; $p < 0.0001$). No significant effect of group ($F_{1,23} = 0.74$; $p = 0.3975$), a significant effect of practice ($F_{3,69} = 15.95$; $p < 0.0001$), but no practice by distance interaction ($F_{9,207} = 1.65$; $p = 0.1019$), distance by group interaction ($F_{3,69} = 0.19$; $p = 0.9055$), or practice by group interaction ($F_{3,69} = 2.01$; $p = 0.1207$).

These results indicate that subjects´ reports of using a spatial image representation to solve these problems are reflected in the question answering profiles. Even in situations of indeterminacy, in which the structure does not allow for a veridical ordering of the items mentioned, subjects will still impose some ordering so that such a representational device can be used to answer the questions.

Premise Integration Time

Premise integration time refers to the time taken for subjects to read the information in all four premises in order to be able to answer the question associated with the premises, and figure 5.2.3 (overleaf) presents the total premise integration time averaged for the four practice blocks for each group of subjects. Analysis of variance on these data reveals an overall effect of group ($F_{2,35} = 3.62$; $p = 0.0373$), an effect of practice blocks ($F_{39,105} = 16.76$; $p < 0.0001$), but no block by group interaction ($F_{6,105} = 1.88$; $p = 0.0904$). This suggests an overall effect of practice on premise integration time.

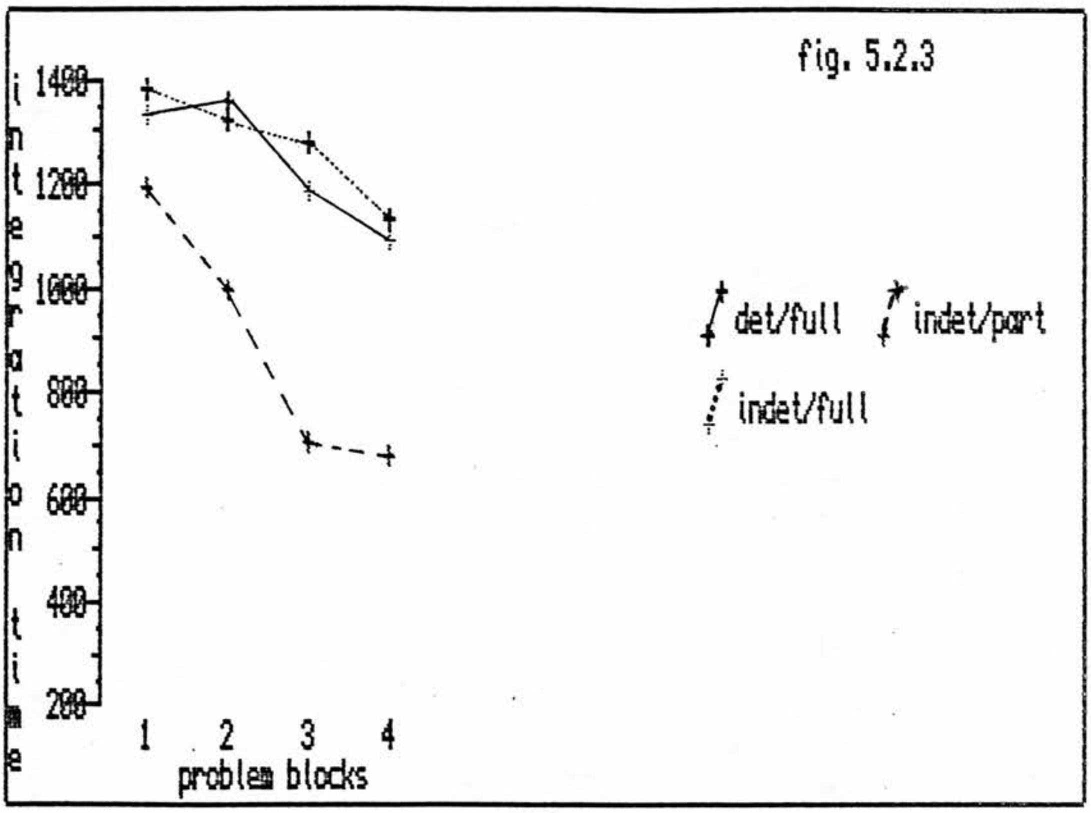


figure 5.2.3 Group comparisons of Premise integration time as a function of problem blocks

Further analyses reveals that the main effect of group is located between group 3 and the other two groups. Thus there was no effect of group when only groups 1 and 2 were considered ($F_{1,24} = 0.07$; $p = 0.7979$), the effect of practice blocks however remained ($F_{3,72} = 7.59$; $p = 0.0002$). This refinement of the group effects allows us to conclude that it is the question set distinction (full vs. partial), rather than the determinacy or indeterminacy of the premises that determines premise integration times.

Figure 5.2.4 a,b&c, (facing page) presents integration times for each of the four premise figures as a function of practice blocks and groups. An analysis of variance on these data reveals a significant effect of figure ($F_{3,105} = 10.35$; $p < 0.0001$), a significant effect of group ($F_{2,35} = 3.62$; $p = 0.0373$), a figure by group interaction ($F_{6,105} = 4.04$; $p = 0.0011$), a figure by block interaction ($F_{9,315} = 2.19$; p 0.0228), but no significant three-way interaction ($F_{18,135} = 2.19$; $p = 0.3087$).

In a consideration of groups 1 and 2 treated separately there was an effect of figure ($F_{3,73} = 12.93$; $p < 0.0001$) but no effect of group ($F_{1,24} = 0.07$; $p = 0.7979$), and no interaction between group and figure ($F_{3,72} = 2.16$; $p = 0.1007$), or between practice and figure ($F_{9,216} = 1.37$; $p = 0.2033$). For group 3 considered alone there was a significant effect of figure ($F_{3,33} = 3.88$; $p = 0.0176$), and practice ($F_{3,33} = 10.08$; $p < 0.0001$), but no interaction between practice and figure ($F_{9,99} = 1.79$; $p = 0.0804$).

These results suggest that the time taken to integrate the premise information varies as a function of premise figure for both partial and full question conditions, and that these effects persist over practice. As can be seen by inspection however the figure profiles for the two question conditions is quite distinct. For the full question condition the rank ordering of problem difficulty seems to be **fig.1 < fig.3 < fig.2 < fig.4**, whereas for the partial condition the ordering is **fig.4 < fig.1 < fig.3 < fig.2**, suggesting that the different question sets place different demands on the information processing carried out by the subject.

On the basis of the protocol reports it is possible to deduce which figures should be expected to produce most difficulty for subjects, particularly in the full question condition. Subjects in these conditions reported the use of a spatial image device in which the preferred direction of working for most subjects was from tall to short. On this basis it would be predicted that figure 1 and 2 would be easier to solve than figures 3 and 4 since they are consistent with subjects' professed direction of working. This seems to be supported by figure 5.2.4. There is a further distinction to be made on the basis of the congruence of the items with respect to their final placement in the array (cf Potts and Sholz). Figure 2 although presenting information in a form consistent with the subjects preferred direction of working, presents new items in a way which is incongruent with the items final placement.

As can be seen from figure 5.2.4, this problem is more difficult to solve than its congruent counterpart- figure 1. Similar profiles emerge for figures 3 and 4 with the congruent figure, ie. figure 3, being solved more easily than figure 4. As can be seen by inspection the size of the effects due to congruity are approximately equal to those due to preferred direction of working. It will be noted that no such profile emerges for the partial question condition.

In a post-test questionnaire administered to subjects they were asked to rank the four premise figures thay had solved in terms of an assessment of their difficulty. And there was a high concordance between these subjective rankings and rankings on the basis of the above predictions from the image model. 80% of subjects using an image strategy record a ranking of fig 1 < fig 3 < fig 2 < fig 4 for example.

In an attempt to analyse these figure effects more fully, individual premise reading times were analysed with respect to figure effects and practice blocks. These results are presented in figure 5.2.5 (facing page). An analysis of variance on these data reveals an overall effect of group ($F_{2,36} = 5.12$; $p = 0.0111$), an effect of practice ($F_{3,108} = 6.97$; $p = 0.0002$), but no main effect of premise ($F_{3,108} = 1.74$; $p = 0.163$). There was no premise by practice interaction ($F_{9,324} = 1.46$; $p = 0.0002$), but there was a significant premise by figure interaction ($F_{99,324} = 2.98$; $p = 0.0002$), and an interaction between group and premise ($F_{6,108} = 2.54$; $p = 0.0246$). Finally there was a significant three-way interaction between practice, figure and premise ($F_{27,927} = 1.59$; $p = 0.0028$), but no interaction between group, block and premise ($F_{18,324} = 1.09$; $p = 0.3652$).

These results suggest that the figure effects obtained when total premise integration time is analysed can be further refined into distinctive premise contours associated with each of the figures. Furthermore, the significant three-way interaction suggests that for some groups, these contour effects change as a function of practice.

An analysis of groups 1 and 2 alone reveals a significant effect of figure ($F_{3,75} = 8.96$; $p = < 0.0001$) and a premise by figure interaction ($F_{9,225} = 3.15$; $p = 0.0013$), but no practice by premise interaction ($F_{9,225} = 1.39$; $p = 0.1925$), and no three-way interaction between block figure and premise ($F_{9,225} = 1.21$; $p = 0.2278$). This suggests that for the full question condition premise contours associated with the various premise figures are stable over practice.

In a consideration of group 3 however, there is not only a main effect of premise ($F_{3,33} = 2.87$; $p = 0.0513$), but a premise by block interaction ($F_{9,99} = 4.46$; $p < 0.0001$), and although there is no interaction between figure and premise ($F_{9,99} = 1.79$; $p = 0.0800$), there is a three-way interaction between premise, block and figure ($F_{27,297} = 1.64$; $p = 0.022$). Thus the partial question condition produces a quite different effect on premise reading time. Here it appears that the premise contours associated with the various figures change as a function of practice.

One of the most striking characteristics of the contour effects is the way in which the two groups solving indeterminate problems - groups 2 and 3 - are almost indistinguishable early in practice but the group with the partial question set quickly drop away from the other group and the premise contour becomes very much flattened.

Indeed there is no significant effect of premise contour for group 3 when only the last two blocks of problems are considered ($F_{3,99} = 0.58$; $p = 0.6351$). This is may reflect the way in which these subjects move away from the image formation strategies as the partial question set becomes exploited by the use of elimination.

Discussion

The hypotheses under test in this experiment are well supported by the data. In consideration of the full perspective question set, once again the protocol reports make it clear that nearly all subjects use some form of image device for solving these problems. These reports are substantiated by observations of inference distance effects in the question answering times. The fact that there was no distinction between subjects solving determinate and indeterminate problems in this full perspective condition suggests that the image device is not constructed as a consequence of premise comprehension but rather is a background resource which is used in order that comprehension might be achieved. This is supported by protocol reports suggesting that practised subjects construct a spatial array in anticipation of reading particular problems and as a hypothesis about those problems.

The protocol reports also make it clear that for the full perspective condition once again, use of the image device persists even as subjects become more expert. Furthermore the characteristic architecture of this device is also stable over practice. Thus even practised subjects claim to have a preference for ordering items along a horizontal axis with the tallest item on the left and preferred left-to-right direction of working.

Having used protocols to establish this stability over practice it was then possible to investigate premise figure effects. In particular it was found that those problems which were incongruent with the subject's preferred direction of working required longer premise integration times and were judged to be more difficult by subjects. These premise profiles are taken as evidence for the existence of information processing strategies which serve to make incongruent information congruent with the characteristic architecture of the underlying representational space. While the underlying representational space is a general cognitive resource, these processing strategies serve as the bridge between this general resource and the specific characteristics of the domain at hand.

These figure effects were present only in the full perspective condition and were stable over practice. In the partial perspective condition, where there was very little evidence of the sustained use of images, these figure effects quickly dissipated as the subject became more expert. This suggests that the processing strategies hypothesised to underlie these effects are indeed dedicted to the image representation and that for an elimination strategy the relationship between premises is of little importance.

The fact that for image users, the premise figure effects are stable over practice is crucial to the understanding of domain specific expertise. This refutes the proposition that domain specific expertise is that which renders formally equivalent problems psychologically equivalent. It would seem that a set of formally equivalent problems are <u>never</u> psychologically equivalent for novice or for expert. Surface structure variation within a class of problems will always make itself felt. The structure of the psychological representation of a problem class is never isomorphic with the formal structure of that class.

The view of domain specific expertise that emerges here then, is that an expert subject is able to recognise a class of problems as potentially solvable with respect to some representational resource. Not only this, but he must also be able to deploy specific procedures for relating this novel class of problems to that background resource.

Given this interpretation, it is clearly of interest to examine how a subject might determine when a particular representational resource is appropriate for a particular problem domain. The clear dichotomy between full and partial problem perspectives obtained in this experiment would suggest that one important criterion used by subjects is a sensitivity towards the functional goals of a problem solving situation. Subjects in the full perspective conditions were required to make multiple comparisons between premise items in order to solve the problem correctly, whereas the partial perspective subjects were only required to identify one item that had a particular property (ie the tallest or smallest item).

The different functional goals of these two conditions make the subject's choice to use an image device more or less appropriate. In the full perspective condition it is an appropriate resource to use. In the partial condition it is not since it involves the subject in retention of premise information which is not pertinent to the solution of the problem. It is not surprising then that subjects in the partial perspective condition abandon the image device in favour of the more appropriate elimination method.

Sensitivity to functional goals was also identified as a factor in the solving of state space problems (cf chapter 3) and is clearly a general problem solving skill probably used by subjects to avoid the use of inappropriate problem solving methods in a wide variety of situations. In a short follow up study (see appendix B). The subjects of this experiment were tested again one week later. This time all subjects were presented with the determinate problems but in a partial perspective condition. Subjects practised in the use of elimination were initially much better at solving these problems despite the fact that group two had experienced these same determinate premises the week earlier. Most subjects who had experienced the full perspective condition the week previously began by using the previously appropriate image device. But this was quickly abandoned however in favour of elimination such that initial group differences were quickly lost. This would suggest that subjects are sensitive to the efficiency of problem solving methods with respect to the kinds of memory load they place on the system for example, as well as the effectiveness of a method in terms of whether it affords a solution to the problem. The image device is an effective method even for the partial perspective condition but it is not efficient.

General Discussion of the Inference Experiments

The decision to use inference tasks rather than state space tasks in this second set of experiments was motivated by two things. Firstly, it was argued that these problems were amenable to protocol analysis techniques and this would provide a relatively direct method of gaining information about the psychological mechanisms underwriting domain specific expertise. Secondly, it was argued that such expertise could be easily studied in this domain of problems because it was possible to systematically manipulate surface structure variation within a single class of problem.

The experiments reported here have certainly born out this decision. In both experiments, but particularly experiment three, protocol analysis has proved to be an affluent source of information concerning the psychological mechanisms of expertise. These protocols revealed that subjects use a form of spatial image device to solve these problems, but contrary to indications in the literature, this device is not abandoned as subjects become more expert. On the contrary, the device is used by expert subjects in anticipation of premise information and as a schema for the interpretation of such information. Furthermore, as the subjects become more expert the device becomes more dimensionalised with subjects reporting a consistent preferred direction of working, along a horizontal axis from left to right with the tallest at the left. In addition, the protocol reports also revealed that with practice the image device became elaborated by the use of mnemonic strategies such as rehearsal, a clear indication that the subject was sensitive to the memorial demands placed on him by these tasks.

Because the protocol reports indicated the image device to be stable over practice, it became clear why surface structure variation plays such an important part in the psychology of reasoning with these inference problems (see also Johnson-Laird 1983). Surface structure variation leads some problems to be presented in a way which is incongruent with the characteristic architecture of the image device. Thus a subject must deploy information processing strategies which render these problems congruent. The need to deploy such strategies with certain forms of problems is responsible for the characteristic premise integration time profile observed with these problems.

It is these processing strategies and the distinction between representation and strategy that they imply, which provides us with an understanding of the psychological factors involved in domain specific expertise. Expertise within a class of structurally equivalent problems does not serve to render such problems psychologically equivalent. On the contrary, domain specific expertise implies that subjects have a quite general representational resource against which novel problems must be evaluated. In so doing the subject must actively translate such novel problems into a form which is compatible with the background resource. Domain specific expertise is the ability to recognise problems as solvable with respect to an existing representational schema while at the same time recognising that each problem has a distinctive surface structure form.

These experiments also confirmed another form of expertise which did not involve the use of the image device. This was the elimination strategy of Wood et al (Wood et al op cit, Quinton and Fellows op cit). It was argued that this expertise was not of the domain specific sort with which this thesis is primarily concerned. This could not be so because it involved the subject in implicitly redefining the problem that he was solving. The shift from the use of the image device to the use of elimination is based upon the subject using his knowledge of functional invariants in the experimental procedure to avoid carrying out inferences altogether.

Based as it is however, on the subject´s awareness of functional goals, this form of expertise can be seen as being rooted in a more general problem solving skill, namely, the subject´s sensitivity to what it is that he ´needs to know´ in order to solve the problem as effortlessly as possible.

It is clear then that the use of protocol reporting has provided information without which a number of the conclusions could not be drawn. For example without an independent means of assessing a subjects preferred direction of working, the effects of surface structure variation would have been uninterpretable. It seems clear that future work in this area must be prepared to consider the role of such data much more carefully.

It is of interest then, to note that there have recently been reports in the literature on the role of verbal reporting in problem solving (Ericsson and Simon 1980, Berry and Broadbent 1984) which have attempted to assess the precise relationship between verbalisable knowledge and problem solving performance. In this thesis we have concentrated very little on this aspect of the protocol data. As was pointed out in the discussion of experiment three, it is not clear from these experiments in what way the observed changes in protocol reports were related to improvement in reaction times. It is not known for example whether practice at protocol reporting per se improved the subjects ability to introspect about his solution method, or whether the changes in problem solving behaviour suggested by the protocols were causal to reduction in solution times.

Berry and Broadbent´s paper serves to demonstrate that the answer to this problem is very complex. They found that the ability of subject to verbalise knowledge associated with performance on complex cognitive tasks was often quite unrelated to the ability of subjects to actually perform the task. Under certain conditions, subjects showed an improvement in performance with practice that was not accompanied by an improved ability to answer questions requiring explication of the knowledge associated with such performance. Under other circumstances, with the very same tasks, subjects would show a marked improvement in their ability to verbalise knowledge but showed no improvement in performance. Only when subjects were given explicit verbal training on the problems and were required to ´think aloud´ while solving the task was there an improvement in perfomance which was accompanied by an improvement in verbalisation.

Clearly, this issue is an important one but one which is in need of further research before an adequate account is forthcoming. From the point of view of this thesis we must be content that the correlation between such measures as premise integration time and inference distance effect, and the qualitative measures of protocol analysis is both consistent across subjects and between experiments, and as such is strongly suggestive of a genuine relationship.

Finally, given that the inference problems were chosen precisely because they supported the use of protocol measures, it might be asked to what extent are such problems representative of problem solving in general? Berry and Broadbent, for example argue that since protocol data does not always provide access to the kinds of knowledge involved in problem solving its importance for a study of problem solving might be quite limited. This conclusion has to be resisted. Apart from the fact that the experiments reported here make claims about the nature of expertise which are in principle generalisable to other domains, there would appear to be strong reasons for suggesting, that only when problem solving involves verbalisable knowledge is it problem solving of any interesting sort. In most areas of 'real world' problem solving such as mathematics and computer science for example, one of the hallmarks of an 'expert' is his ability to communicate his understanding of the subject to others. If this were not the case then science would simply not be possible. It is also of interest to note in this respect, that one of the criteria used in AI research to assess the 'expertness' of an expert system is the ability of the system to explain to the user the reasoning behind its solution.

CHAPTER SIX

CONCLUSIONS

An Evaluation of the Experimental Programme

This thesis has been concerned with an investigation into the nature of expertise in problem solving. In the introductory chapter a number of criteria were put forward against which much of the literature on problem solving could be assessed.

It was argued that a distinction had to be maintained between a ´successful´ solution and an ´expert´ solution. The latter is not a solution which simply provides an answer to a problem but one which is tempered by considerations of economy, and particularly one which extends beyond the confines of the problem in which it was generated. It follows from this latter point, that any experimental study of expertise must use transfer of learning paradigms in some form or other. Much of the early research on animal learning was judged to be inadequate with respect to this criterion (Cf Thorndike 1898, Kohler 1925 for example). Few of these studies used transfer paradigms, and those that did used the train-and-test paradigm simply as a means of testing for the presence of a particular solution method.

It was argued that a study of expertise required a different kind of transfer paradigm, one in which subjects were presented with multiple problems to solve and which allowed for a level of expertise to be generated within the testing situation. While there had been such studies reported (Cf. Harlow 1959, Hull 1920), these studies proved to be inadequate by a further criterion.

It was argued that there were two possible views about the nature of expertise. The first of these regards expertise as the ability to use highly effective problem solving methods but methods which are quite task specific because they are based on knowledge which is specific to the task domain in question. This domain specifc notion of expertise was contrasted with a second which views expertise more in terms of the possession of problem solving heuristics. These heuristics are not tied to a particular task domain but are quite general because they do not incorporate task-specific knowledge. While studies using multiple-problem transfer procedures had demonstrated quite dramatic transfer effects they had failed to discover which kind of expertise was underwriting such transfer effects.

Because these criteria proved so useful for evaluating the existing problem solving literature, it seems appropriate at this time to assess the experiments reported in this thesis by the same criteria. The first goal of the experimental programme was to establish transfer effects in the laboratory and to identify which form of expertise might underwrite any transfer thus obtained. The state space problems of Luger et al (Goldin and McClintock 1980) were useful in this respect since they allowed for a means of analysing problems into classes or domains in terms of their structural affinities. Furthermore it had been argued that human problem solvers are sensitive to the structural properties of these problems and that knowledge of structural equivalence was involved in the solution of these problems. Thus if expertise in problem solving were of the domain specific sort, it would be expected that any transfer effects that were obtained with these problems would be limited to problems of the same structural type. In contrast if expertise were of a more general sort then no such constraints on transfer would be expected.

The results of the two experiments with state-space problems reported in this thesis show that in fact both kinds of transfer effects can be obtained in the laboratory. Experiment 1 revealed transfer effects across non-isomorphic problems. This was taken to indicate that some form of non domain-specific expertise had been tapped. In contrast, experiment 2 revealed transfer effects which were limited to isomorphic problems and thus taken to indicate that some form of domain-specific expertise had been tapped. This second form of expertise however, produced much more dramatic transfer effects than the more general sort obtained in experiment 1. Here subjects were showing 100% transfer after solving only five problems.

These experiments also made a clear distinction between a 'successful' solution to a problem and an 'expert' solution. In experiment 1 in which subjects were simply required to solve each problem in any way they liked, the solutions tended to be inefficient as judged by the number of redundant moves in the solution path and there was no evidence for any transfer across isomorphic problems. Only in experiment 2 in which subjects were required to solve each problem to a criterion of mastery did they use knowledge of structural equivalence to produce solutions in the minimum number of moves possible, and only then did the dramatic isomorphic transfer ensue. It was clear that expertise within a single problem was necessary to guarantee transfer to subsequent problems of the same type.

While the experiments with the state space problems provided us with the ground rules for identifying expertise of a domain specific sort and made it clear that such expertise produced by far the most dramatic transfer effects, in fact these experiments afforded little that could be used to establish a psychological account of such expertise. It was not clear from these experiments what the nature of the psychological representation might be that allowed subjects to extend solution procedures to problems differing quite widely in their surface structure characteristics. Nor was it clear what the role of practice was on these representational mechanisms. Part of the problem here lay in the fact that the psychological measures that could be used in this problem domain were quite restricted. Variables such as solution time, moves to solution and trials-to-criterion while offering good indices of improvement with practice do not themselves inform about the mechanisms giving rise to such improvements.

In an attempt to rectify these measurement problems the method of protocol reporting was used. Such methods have been shown to be of use in other problem solving contexts and have provided useful information about the nature of the psychological representations that underwrite performance, Furthermore, research using state space tasks has demonstrated the efficacy of this method for providing information concerning the kinds of moves subjects make or consider making.

The attempts in experiment 2 to use this method to obtain information about the representational device underwriting such move-making were unrevealing. Subjects appeared to be unable to report anything about their representation of the problem environment. Since the method of protocol reporting has been shown to be successful in other problem solving situations it could only be concluded that there was something inherent in the nature of these state space tasks that made the method inappropriate. It might be for example that the level of representation is too low to be amenable to conscious access. This explanation has been put forward to account for the fact that people are very poor at describing motoric skills which they demonstrably possess.

A second limitation of the state space paradigm as an empirical basis for psychological explanations is that the problem types are limited in terms of the number of distinct problems that can be constructed for each type, and the individual problems within each type cannot be varied in any systematic way. Clearly, if we are to study factors affecting expertise within a single task domain then the internal structure of that domain must be well specified.

For these reasons it was decided to look for a new problem domain which, while affording potential for investigating some of the psychological mechanisms of expertise, nevertheless conserved the important findings of the previous work. In particular some means of classifying problems in terms of their structural characteristics would be needed. This would allow further experiments to focus on the kind of domain specific expertise that had produced such dramatic transfer with the state space problems. Furthermore it would be necessary for large classes of problems to be generated so that multiple problem transfer procedures could be used. But also it must be possible to specify the systematic ways in which problems within a given class differ from one another. Finally it must be possible to measure transfer in terms of solution time or some other index as well using protocol methods to get at some of the psychological mechanisms that underwrite performance.

With these criteria in mind deductive inference tasks were used as the problem domain for further experiments. The formal logical analysis of these problems provides a taxonomy of problem classes and there was already a wealth of litrature on the use of reaction time data and protocol methods with these problems.

But the work that was already in existence on expertise factors in the solving of these problems was both sparse and ill-conceived. It had been argued that ´novice´ subjects solved these problems by integrating the premise information into some form of spatial image representation or mental model, while ´expert´ subjects avoid the use of any such ´thinking´ strategies. Rather they used short-cut heuristics based on their knowledge of the questions asked and the syntactic invariances in the premises of the problems, (Johnson-Laird 1975, Quinton and Fellows 1975, Wood 1974). But it was clear however that this version of expertise was not of the same nature as that which had been identified in the state space experiments.

This expertise was not generated over a single class of problems. It could not be because the subject was implicitly redefining the class of problems within which he was working. At the outset the subject´s task was to make transitive inferences based on the co-ordination of premise information, but the subject becomes an expert, not at doing this, but rather at searching the premises in turn for a specific item of information. These expert strategies actually allow the subject to avoid making inferences altogether. It was clear then that before we could begin to examine the psychological mechanisms of expertise in these tasks we must establish the existence of a form of expertise that was specific to the inference problems as a distinct class of problems.

In experiment 4 of this thesis a direct comparison was made between subjects solving problems by the use of shallow heuristics, and subjects solving problems that required inferential skills. The results clearly showed that there were significant transfer effects as indexed by reduction in solution time with practice and that it was shown by both groups of subjects. Only in one group however did the protocol reports show the use of shallow heuristics. In the other groups, subjects report using a spatial image strategy. Thus there is demonstrably an expertise factor in performance on these inference tasks which is a true form of expertise on a single class of problems analogous to that obtained in the state space experiments.

Having established that there is some form of domain specific expertise to be explained here, it was then possible to consider the extent to which the protocol data offer some insights into the psychological nature of the expertise and how this is expressed within the problem domain, thus justifying the decision to use inference problems. Experiment 3 used a protocol probe technique to analyse the characteristics of the subjects problem solving strategy and how this changes over practice. Most other studies have simply asked subjects to provide protocol reports after they had finished solving the problems, in this study subjects were interrupted at various points in the experimental session and asked to provide a protocol report. This method proved to be an affluent source of information as to the psychological variables affecting performance.

Firstly it was found that nearly all subjects used some form of spatial image device for representing the items mentioned in the premises in terms of an ordered linear array. The use of such devices has been reported elsewhere in the literature but the status of such a device with respect to expertise was unknown. It has been argued that such a strategy was only used by novice subjects, on the other hand the developmental literature suggests that the use of such a device is a quite sophisticated response to the problem.

The results of the protocol reports make it quite clear in the case of adult problem solvers solving problems which do not support the use of shallow heuristics, that the image device occurs very early in the path-to-expertise. Even after the very first problem, most subjects report using such a device, furthermore the use of this device persists despite improvements in performance as their experience with the problem domain grows. Most subjects still reported using the device after the last problem of the session.

It is clear then that the use of this device is not a consequence of growing expertise. The results of experiment 4 add further weight to this. Here it was found that subjects report using the device quite consistently, even for solving indeterminate problems which do not in themselves provide enough information for the construction of such an image. This made it clear that subjects were not constructing the image on the basis of an understanding of the premise information but rather using the image as a schema for the interpretation of that information. This was further supported by the protocol reports of subjects using the spatial image device as a hypothesis in anticipation of actually reading the premises of a novel problem.

Thus rather than being a consequence of expertise the spatial image device appears to be a general background resource capable of being called up in response to a diverse range of problem solving situations. It is hardly surprising then to find that the use of such devices has been widely reported in many other contexts (Lakoff and Johnson 1980, Foos 1980, Friedman 1984), and indeed that subjects in the experiments reported here can not only use such devices to solve three- and five-term transitivity problems but also indeterminate problems and problems involving the co-ordination of two-dimensional relational terms.

But if the spatial image device is to be interpreted as a general problem solving resource, how is the expertise factor tapped in these experiments to be explained? Once again the protocol reports from experiment three provide some clues. These indicate that although the use of a spatial image device persists throughout all the problems the basic mechanism is elaborated somewhat as a function of practice. One significant elaboration is the use of mnemonics such as rehearsal. But while such mnemonics provide interesting information concerning metacognitive factors in the information processing abilities of these subjects, they may have little to do with the expertise factor observed here. On the contrary it could be argued that such mnemonics emerge as a response to the extra time for which expert subjects were required to retain premise information .

Of more interest were the reports suggesting that the image device became more clearly dimensionalised as a function of practice with the items being arranged on a horizontal vector from left to right with the tallest on the right. Subjects began to actively search for premise information which was congruent with this preferred direction of working. It seemed that as suggested by McGonigle and Chalmers (1984 a&b), there are two aspects to the solving of these problems that must be differentiated. Firstly there is the representational space that is prior to the interpretation of any problem, and which has a characteristic architecture, and was used as a schema for interpreting premise information. Secondly there must be a set of information processing strategies that subjects use to map the different surface forms of the problem domain onto this space. Problems whose surface forms were most congruent with the underlying representational architecture require the least extra processing to be so mapped, but problems whose surface forms are incongruent with the representational space require extra processing time to make them congruent. In terms of a task specific component of expertise then, it seemed that this could be identified as the ability to develop strategies that allowed the problems of a specific domain to be interpreted in terms of a more general, domain-independent representational resource.

Experiment 4 examined this hypothesis further by presenting subjects with 4 sets of problems which, although formally equivalent five-term inference problems, differed markedly in the degree to which there surface forms were congruent with the known characteristics of the spatial image device. It was argued that incongruent versions would require extra processing time for their premises to be mapped onto the representational space. But as the subject´s knowledge of the problem domain grows, and he becomes more expert at dealing with the specific surface structure variations we might expect that these differences between congruent and incongruent forms would dissipate. This would be quite consistent with the operational definition of expertise that was developed in response to the state space findings namely, expertise is that which is required in order to guarantee that formally equivalent tasks are also psychologically equivalent.

The premise integration time data of experiment 4 provided strong support for the claim that there was a meaningful distinction to be made between representation and strategy. There were strong effects of problem congruence on premise integration time. Problems that were less congruent took longer to integrate. However the effect of problem congruence did not interact with practice effects, incongruent problems did not become easy to integrate as subjects became more expert. Indeed if anything the effects of congruence became more stable with practice.

It appears then that these findings contradict the hypothesis that while the representational space is a general problem solving resource the processing strategies constitute a domain specific factor of expertise. But more careful consideration allows this contradiction to be resolved. If we accept the good evidence to show that the representational architecture is independant of the particular information that is presented in a problem and remains unchanged as a result of that information, then obviously the processing strategies that are used to map that information onto the representational space must also be stable over practice. Indeed it might even be expected that as the problem domain becomes more well defined with practice then the reaction time profiles attributable to the use of such strategies must themselves become more well defined. Any changes in such profiles could only be a consequence of a general noise reduction factor and would perhaps be of little interest.

To further clarify this we might contrast the situation in experiment 4 with that in experiment 3. In experiment 3 subjects report scanning problems in search of premises that were congruent with their preferred direction of working. Such scanning was obviously not possible in experiment 4 because premises were presented individually. But if subjects in experiment 4 had been able to control the order of premise presentation, it might have been found that, as their experience of the surface variations present in the problem domain grew, they might well have begun to opt for a presentation order that meant premises were presented in an order that was congruent with their preferred direction of working. This would have alleviated the need for inter-premise conversion strategies and so reduced premise integration time. But this would not have meant that the effects of problem congruence had been dissipated in any true sense, it would merely have meant that the incongruences had been overcome in a different way.

Whichever way it is viewed then, domain specific expertise is the ability to deploy information processing strategies that allow novel problems in that domain to be interpreted in terms of a more general representational framework. But this interpretation of domain specific expertise contradicts the earlier suggestion that the effect of such expertise was to render a formally equivalent class of problems psychological equivalent. The effects of surface structure variation on 'naive' problem solvers is well documented in both the developmental and adult literature (Donaldson 1968, Wason and Johnson-Laird 1972 for eg), and it might have been expected that extended training on structurally equivalent problems would serve to alleviate the adverse effects of such content variables on abstract thinking. But the conclusions of this thesis make it clear that formally equivalent problems, varying in their surface forms, are never psychologically equivalent for novice or expert. Indeed they will be treated as different by an expert precisely because they are formally equivalent. The only difference between novice and expert is that the expert, if he knows two problems to be formally equivalent, can attempt to provide a solution for one on the basis of his solution for the other, but this involves knowing the systematic differences between the two versions of the problem. This distinction between formal and psychological equivalence is reflected clearly in a recent article on the solving of 'NP-complete' polynomial equations in mathematics. After much research effort these problems are still refractory to solution. More recent research effort has been dedicated instead to proving that these problems form a class and are related to other classes of problems for which solutions might be more readily attained. In which case mathematicians would still not have a solution to the NP-complete problems, but at least they would know that a solution exists (Kolata 1980).

Directions for further research

To summarise then, the experimental programme reported in this thesis has achieved several things. Firstly it has shown that expertise can be studied in the laboratory, provided that suitable experimental procedures are employed. Secondly, it has distinguished between two forms of expertise and provided some evidence for both. Finally in concentrating on one form of expertise namely, domain specific expertise, the thesis has offered some account of the psychological mechanisms involved. In considering the implications this work has for further research, there would seem to be two useful directions further work might take. The first of these is to examine some of the psychological mechanisms that are involved in the other form of expertise observed, but not accounted for, in this thesis, namely non domain-specific expertise. The second is to follow up the major conclusion of the thesis with respect to domain specific expertise, namely to examine further the relationship between general representational resources and task specific processsing strategies.

In considering non domain-specific expertise, the results of the thesis suggest that metacognitive factors play an important role in problem solving and would seem to offer a strong candidate for aspects of some general problem solving skill. The factors can be classified into two types, which might be labelled, 'subject' and 'object' knowledge. (cf. McGonigle 1984). Sensitivity of subjects to their own memory limitations is an example of subject knowledge. The protocol reports revealed that subjects were very much aware of their working memory limitations as they solved the inference tasks. They were able to predict when characteristics of the task would be likely to lead to failures of memory and were able to take pre-emptive action to avoid such failures by the use of mnemonics. This awareness presupposes that human subjects have access to the workings of their own system and knowledge of their own design limitations. Such knowledge would clearly be of use in a wide range of problem solving situations and is thus a strong candidate for a general problem solving skill.

It is interesting to note in this respect that the growth in metacognitive awareness and self-monitoring of this sort has recently been put forward as a major axis of cognitive development (Kail 1978). Young children can be taught to use mnemonics such as rehearsal, but it is not until age 9 and over that they are able to recognise situations in which the use of such mnemonics is appropriate. This distinction between ´knowing how´ and ´knowing when´ has also been identified as an important dimension in children´s oral communication skills. (Dickson 1980).

Object knowedge refers to the knowledge that a subject gains about the functional criteria at work in the problem solving environment. In the case of the inference experiments for example, the shift of subjects towards shallow processing heuristics was based on the subjects´ knowledge that they were only being interrogated about a subset of the information that was available and could thereby eliminate redundant information. Such sensitivity to what it is the subject ´needs to know´, allows highly cost-effective processing of information. Given the difficulty that subjects experience in finding minimum solution paths to the state space problems for example, it would be wasteful for a subject to process to this level if a minimum path was not a requirement of the task as functionally defined by the experimenter. The ability of a subject to assess the functional goal of a novel problem situation is thus a further general problem solving skill.

Research on metacognitive factors is concerned, in essence, with the study of the problem solver as a 'knowing system'. It follows then, that the paradigm for such research must involve informing the subject, quite explicitly, about the nature of the functional goals that are at work in the problem solving situation, and assessing the extent to which such knowledge affects problem solving behaviour. It is of interest to speculate then, whether subjects solving state space tasks for example, would behave differently if they were explicitly informed about the isomorphic structure of problem classes and that mastery of one guaranteed mastery of all. Would subjects, knowing that they had a large number of isomorphic problems to solve, elect to practice extensively on one problem before moving on to the next? Such experiments would raise interesting questions concerning the ability of subjects to assess the computational costs and benefits accruing to various problem solving strategies, and to decide which strategy would be most appropriate in a given situation.

The problem of strategic decision making has received very scant attention in recent years. The emphasis has tended to be on describing the different strategies that different subjects use and explaining individual differences in these terms (Newell and Simon

1972, Egan and Grimes Farrow 1982, Underwood 1978 for eg). There has been little corresponding work on why a given subject chooses the strategy he does. Most of the AI work in this area has accounted for such decision making by recourse to deterministic procedures such as using the strategy that has most recently proved successful and so on (Anderson 1976). But recent work on hypothesis evaluation (Fischoff and Beth-Mayroum 1983) and judgemental heuristics (Kahneman and Tversky 1983), make it clear that human subjects have a much richer set of decision criteria available to them and much work remains to be done in describing exactly what criteria human subjects use to determine the efficacy of a given problem solving strategy in new situations.

Turning now to a consideration of further work on domain specific expertise, the central issue is one of further investigating the relationship between representation and strategy. One of the most important aspects here is the suggestion that the representational space features as background resource used to interpret novel problems rather than being a product of that interpretation. This has important implications for the role of such devices in tasks such as discourse understanding which is discussed by McGonigle and Chalmers (1984b).

From the perspective of the experiments reported in this thesis however, a number of directions for further work are suggested. The most important of these is the need to examine what constraints there are on the use of representational spaces such as the spatial image device. That is to say, what is the domain of problems for which such a device might serve as a background resource?

The experiments reported here have already shown that subjects appear to be able to use this device to interpret a diverse range of problems, from those involving both spatial and non-spatial relational terms in one and two dimensions through to indeterminate problems of certain sorts. The indeterminate problems are a particularly interesting case because there is nothing inherent in the premise information that gives warrant to the use of such a device. But one factor that does give such warrant is the fact that a subject can usually be assured that, although the information necessary for making certain comparisons is not in fact present in the premises, nevertheless, the items that are referred to by the premises, that is the referential domain of the problem, do in principle contain that information. So for example although the premises; "A is taller than C, C is shorter than B; Is A taller than B?" do not contain information about the A-B relation, the subject can be assured that if A, B, C were picked out in the real world they would be orderable along the dimension of tallness. This assurance validates the subject's attempt to impose a determinate ordering on the information.

Perhaps it is the case then, that the use of a spatial image device is constrained by the referential domain of the problem. Consider for example the problem "A is taller than B, B is a better violinist than C; Is A a better violinist than C?". It is not obvious that a subject would be able to solve such a problem by the use of a spatial image device or any other device for that matter. Recent work at Edinburgh has extended the inference studies to look at the relationship between the referential domain of problems and the use of spatial images. They find that when the problems describe items in a way that is congruent with their appearence in everyday contexts then they are comprehended more easily, and secondly thay found that the continuity of the referential domain was a more significant factor in comprehension than the determinacy of the problems.

If it proved possible to delimit the domain of problems for which the image device serves as a background resource, it would then be possible to examine in more detail the range of information processing strategies that a subject is capable of deploying in his attempts to deal with novel problems in terms of this background resource. The experiments reported in this thesis did not discuss in any detail the strategies that subjects use to process indeterminate information determinately and further study of indeterminacy will be revealing here.

In conclusion then, this thesis has laid out the ground rules for an adequate theory of expertise in problem solving and attempted to examine just a small part of the psychological mechanisms that are involved in such expertise, but in so doing, it is hoped that directions for further work have been well specified. One thing that is certain is that there are insights to be had into the nature of problem solving expertise, and that these might only be had by recourse to psychological investigation.

<u>REFERENCES</u>

Adams, J.A. 1954

Multiple Versus Single Problem Training in Human Problem Solving
Jou. of Exp. Psych. Vol. 48 pp. 15-19

Anderson, J.R. 1976

Language Memory and Thought
Hillsdale,N.J: Lawrence Erlbaum Associates

Anzai,Y. and Simon, H. 1979

A Theory of Learning by Doing
Psych Rev Vol 86.1

Berry, D.C. and Broadbent, D.E. 1984

On the Relationship between Task Performance and Associated Verbal Knowledge
Qua. Jou. of Exp. Psych. Vol. 36A.2 p. 209

Breslow, L. 1981

Re-evaluation of the Literature on the Development of Transitive Inference
Psych. Bull. Vol. 89.2 pp. 325-351

Bruner, J.S., Goodnow, J.J. and Austin, G.A. 1956

A Study of Thinking
New York: Wiley

Bryant, P.E. and Trabasso, T. 1971

Transitive Inferences and Memory in Young Children
Nature Vol. 232 pp. 456-458

Bundy, A. 1978

Artificial Intelligence: Introductory Notes
Edinburgh University Press

Callantine, M.F. and Warren, J.M. 1955

Learning Sets in Human Concept Formation
Psych. Repts. Vol 1 pp. 363-367

Clark, H.H   1969

Linguistic Processes in Deductive Reasoning
Psych. Rev. Vol. 76 pp. 387-404

DeSoto, C.B., London, M., and Handel, S.   1965

Social Reasoning and Spatial Paralogic
Jou. of Pers. and Soc. Psych. Vol. 2 pp. 513-521

Dickson,   1980

Children's Oral Communication Skills

Diennes, Z.P. and Jeeves, M.A.   1970

The Effects of Structural Relations on Transfer
Hutchinson Educational Ltd.

Donaldson, M.   1968

Children's Minds
London: Fontana

Egan, D.E. and Grimes-Farrow, D.D.   1982

Individual Differences in Strategies for Reasoning
Mem. and Cog. Vol. 10.4 pp. 297-307

Ericsson, K.A. and Simon, H.A.   1980

Verbal Reports as Data
Psych. Rev. Vol. 87.3

Ernst, G. and Newell, A.   1969

GPS: A Case Study in Generality and Problem Solving
New York: Academic

Falmagne, R.J. (Ed.)   1975

Reasoning: Representation and Process in Children and Adults
Hillsdale: Erlbaum

Fischoff, B. and Beyth-Marom, R.   1983

Hypothesis Evaluation from a Bayesian Perspective
Psych. Rev. Vol. 90.3 pp. 239-260

Foos, P.W. 1980

Constructing Cogntive Maps from Sentences
Qua. Jou. of Exp Psych. Vol. 6.1 pp. 25-38

Friedman, W.J. 1984

Reasoning about Months of the Year
Jou. Exp. Psych. Vol 9.4 pp. 650-660

Goldin, G.A. and McClintock, C.E. 1980

Task Structure Variables in Mathematical Problem Solving
ERIC, Ohio State University

Greeno, J.G. 1974

Hobbits and Orcs: Acquisition of a Sequential Concept
Cog. Psych. Vol. 6 pp. 270-292

Grose, R. and Birney 1963

Transfer of Learning
Van Nostrand

Harlow, H.F. 1949

The Formation of Learning Sets
Psych. Rev, Vol. 56 pp. 51-65

Harlow, H.F. and Warren, J.M. 1952

Formation and Transfer of Discrimination Learning Sets
Jou. of Comp. and Physiol. Psych. Vol. 45 pp. 482-489

Harlow, H.F. and Levine, M. 1959

Learning Sets with 1- and 12- trial Oddity Problems
Am. Jou. of Psych. Vol. 72 pp. 253-257

Harlow, H.F. 1959

Learning Set and Error Factor Theory
in Koch, S. (Ed.) pp. 492-537

Hayes J.R. and Simon, H.A. 1976

Psychological Differences among Problem Isomorphs
in Potts, G.

Hull, C.L. 1920

Quantitative Aspects of the Evolution of Concepts
Psych. Mono. Vol. 28

Hunter, I.M.L. 1957

The Solving of the 3-Term Series Problem
Brit. Jou. of Psych. Vol. 48 pp. 286-298

Huttenlocher, J. 1968

Constructing Spatial Images: A Strategy in Reasoning
Psych. Rev. Vo. 75 pp. 550-560

Johnson-Laird, P.N., Legrenzi, p. and Legrenzi, M.S. 1972

Reasoning and a ense of Reality
Br. Jou. of Psych. Vol. 63.3 pp. 395-400

Johnson-Laird, P.N. 1975

The Three-Term Series Problem
Cog. Vol. 1.1 pp. 57-82

Johnson-Laird, P.N. and Wason, P.C. 1977

Thinking: Readings in Cognitive Science
Cambridge University Press

Johnson-Laird, P.N 1982

Thinking as a Skill
Qua. Jou. of Exp. Psych. Vol 34a pp.1-29

Johnson-Laird, P.N. 1983

Mental Models
Cambridge University Press

Jones, S. 1970

Visual and Verbal Processes in Problem Solving
Cog. Psych. Vol. 1 pp. 201-214

Kahneman, D. and Tversky, A. 1983

Extensional Vs. Intuitive Reasoning:
Psych. Rev. Vol. 90.4

Kail, 1978

Metamemory
Routledge Kagan Paul

Kendler, T.S. 1961

Concept Formation
Ann. Rev. of Psych. Vol. 12 pp.447-472

Koch, S. 1959

Psychology: A Study of a Science
Vol. 2 New York: McGraw Hill

Kohler, W. 1925

The Mentality of Apes
London: Kegan Paul, Trench, Trubner and Co.

Kolata, G.B. 1980

Solve One And You Could Solve Them All
New Scientist Vol. 86 pp. 10-12

Lakoff, G. and Johnson, M. 1980

Metaphors We Live By
Univ. Chicago Press

Levine, M., Levinson, B., and Harlow, H.F. 1959

Trials Per Problem as a Variable in Learning set
Jou. Com. Physiol. Psych. Vol. 52 pp. 396-398

Luger, G.F 1976

The State Space and Behavioural Effects in the
ToH problem.
Int. Jou. of Man-Machine Studies vol 8 pp 411-421

Luger, G.F. and Bauer, M.A. 1976

Transfer Effects in Isomorphic Problem Situations
Acta Psych. Vol. 42 pp. 121-131

Luger, G.F 1978

A State-Space Description of Transfer Effects
Int. Jou. Man-Machine Studies Vol. 10 pp.613-623

Luger, G.F. 1980

State-Sapce Representation of Problem Solving Behaviour
in Goldin, G.A. and McClintock, C.E.

Luger, G.F. and Steen, M.   1981

The State Space and Symmetry
Int. Jou. Man-Machine Studies Vol. 14 pp. 449-460

Mackintosh, N.J.   1974

The Psychology of Animal Learning
Academic Press

Mani, K. and Johnson-Laird, P.N.   1982

The Mental Representation of Spatial Descriptions
Mem. and Cog. Vol. 10.2 pp. 181-187

McGonigle, B.O. and Chalmers, M.A.   1977

Are Monkeys Logical?
Nature Vol. 267 pp. 695-696

McGonigle, B.O and Jones, B.T.   1978

Levels of Stimulus Processing by the Squirrel Monkey
Perception Vol. 7 pp. 635-659

McGonigle, B.O.   1984

The Evolution of Intelligence from the
Standpoint of Comparative Psychology
Proceedings of Tyson Foundation conference

McGonigle, B.O. and Chalmers, M.A.   1984a

Are Young Children Any More Logical the Monkeys?
Jou. of Exp. Child Psych.

McGonigle, B.O. and Chalmers M.A.   1984b

An Information Processing Approach to Inference
in Myers T.K., Brown, E.K., and McGonigle, B.O.

Morrisett, L. and Hovland, C.I.   1959

A Comparison of Three Varieties of Training in
Human Problem Solving
Jou. of Exp. Psych. Vol. 58 No. 1

Myers, T. Brown, E.K., and McGonigle, B.O. (Eds)   1984

Reasoning and Discourse
Academic Press

Mynatt, B.T. and Smith, K.H. 1977

Constructive Processes in Linear Order Problems
Revealed by Sentence Study Times
Jou. of Exp. Psych. Vol. 3, No. 4 pp. 357-375

Newell, A. and Simon, H.A 1972

Human Problem Solving
New Jersey; Prentice Hall

Nilsson, N. 1972

Problem Solving Methods in Artificial Intelligence
New York: McGraw Hill

Ohlsson, S. 1980

A Possible Path to Expertise in the 3-term Series Problem
Cognitive Seminar Working Paper, Univ. of Stockolm

Potts, G.R. and Sholz, K.W. 1975

The Internal Representation of a 3-term Series Problem
Jou. of Verb. Learn. and Verb. Behav. Vol. 14 pp. 439-452

Potts, G. (Ed.) 1976

Indiana Cognitive Symposium
Potomac Maryland: Lawrence Erlbaum Assoc.

Quinton, G. and Fellows, B.J. 1975

Perceptual Strategies in the solving of
Three-Term Series Problems
Brit, Jou. of Psych. Vol. 66 pp. 69-78

Reed, S.K., Ernst, G.W. and Banerjee, R.B. 1974

The Role of Analogy in Transfer between Similar
Problem States
Cog. Psych. vol. 8 p.436

Reese, H.N. 1968

Perception of Stimulus Relations
New York: Academic

Riley, D.A. 1968

Discrimination Learning
Boston: Allyn and Bacon

Schulz, R.W. 1963

Problem Solving Behaviour and Transfer
in Grose and Birney (eds)

Shusterman, R.J. 1962

Transfer Effects of Successive Discrimination
Reversal Training
Science Vol. 137 pp. 422-423

Simon, H.A. 1975

The Functional Equivalence of Problem Solving Skills
Cog. Psych. Vol. 7 p. 268

Simon, H.A. 1980

Problem Solving and Education
in Tuma, D.T. and Reif, F.

Sternberg, R. 1980

The Development of Linear Syllogistic Reasoning
Jou. of Exp. Child Psych. Vol. 29 pp. 340-356

Thomas, J.C. 1974

An Analysis of Behaviour in the Hobbits-Orcs Problem
Cog. Psych. Vol. 6 pp. 257-269

Thorndike, E.L. 1898

Animal Intelligence
Psych. Rev. Mono. No. 8

Trabasso, T., Riley, C. and Wilson, E. 1975

The Representation of Linear and
Spatial Strategies in Reasoning
in Falmagne, R.J.

Tuma, D.T. and Reif, F. 1980

Problem solving and Education: Issues in
Teaching and Research
Lawrence Erlbaum

Underwood, G. 1978

Strategies for Information Processing
Academic Press

Wason, P.C. and Johnson-Laird, P.N. 1968

Thinking and Reasoning
Penguin Books

Wason, P.C. snd Johnson-Laird, P.N. 1972

Psychology of Reasoning: Structure and Content
London: Batsford

Wertheimer, M. 1945

Productive Thinking
New York: Harper

Wood, D.J., Shotter, J. and Godden, D. 1974

Problem Solving Strategies, Representation and Memory
Qua. Jou. of Exp. Psych. Vol. 26 pp. 252-257

Wood, D.J. 1978

The Nature and Development of Strategies
Ch. 10 in Underwood, G.

Yerkes, R.M and Yerkes, A.W. 1925

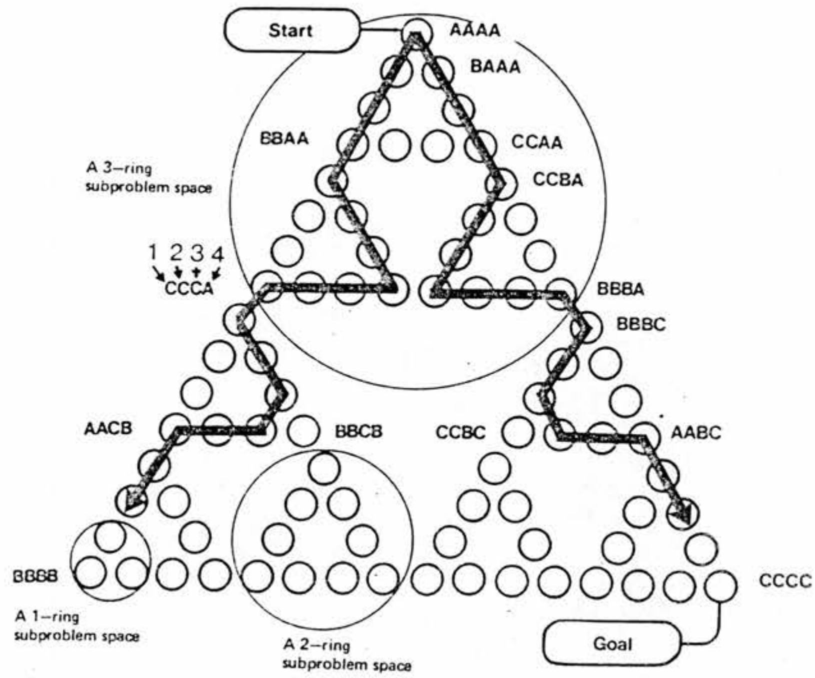
The Great Apes: A study of Anthropoid Life
Yale University Press

APPENDIX A

GRAPHS AND DESCRIPTIONS
OF THE STATE SPACE PROBLEMS
REFFERRED TO IN THE THESIS

APPENDIX A

The State Space For The Tower Of Hanoi Type Problems



Tower Of Hanoi Version 1: The Wedding Cakes

The four tiers of the wedding cake were iced on the table upon which they now stand, that is the red table. For the wedding they must be moved to the blue table.

Because of its weight you must only move one tier at once. Also of course it would be unwise to rest a larger tier on a smaller tier (for example the base on the top tier), since there might be a danger of collapse. It would also be very risky to try and place two tiers side-by-side upon any table as they might topple off the table.

as they might topple off the table.

Despite these problems can you show how to transfer the cake from the red table to the blue table?

ToH version 2: Parking Lots

There are four lorries parked on **road** A on the board in front of you. As you can see the roads are narrow and no overtaking is possible. The lorries are of different sizes and the biggest is at the front and the smallest is at the back and so on. They are all heading for the factory in road C. Outside the factory are four parking lots. Because of the size difference each lorry has its own ´company lot´. There are similar company owned lots in roads A and B. Each lorry may park **only** on the lot appropriate to its size. Can you demonstrate the means by which each lorry may deliver its goods without violating the rules. Please move one lorry at a time.

ToH Version 3: Computer Problems

Imagine I have a very old and cronky computer. My computer has three memory locations. These three memories are represented by the three ´pigeon-holes´ in front of you. Each memory can hold only four numbers. As you can see the top memory has the four numbers 1, 2, 3, 4 in it.

The operating characteristics of this computer are such that it can move numbers from one memory to the other, but can only move one number at once.

Furthermore the computer is so old that no number can be placed on top of a smaller number or else the smaller number will be lost. As you can see though it is alright to put smaller numbers on top of larger ones. If you were the computer programmer what sequence of moves would you instruct the computer to carry out in order to transfer the four numbers in the top memory to the middle memory without losing any of them.

ToH Version 4: Rocket Silos

There are three bays in the rocket factory, the construction bay, the launch bay and the hold bay.
The construction bay contains a brand new space ship. The space ship is made up of four stages or modules as follows;

the Booster	(B)
the Orbiter	(O)
the Mother Ship	(M)
the landing craft	(L)

As can be seen from the board in front of you the landing craft (L) is the smallest and is stored inside the mother ship (M). The mother ship is in turn inside the orbiter (O). The booster (B) is the largest and contains the other three.

The engineers must move the rocket from its present position in the construction bay to the launch bay. They have only one crane and it can lift only one module at once. Furthermore for security reasons no module may stand anywhere except one of the three bays and modules must be stored appropriately where necessary, that is inside one another.

If you were the engineers what sequence of moves would you carry out in order to transfer the rocket from the construction bay to the launch bay.

ToH Version 5: The Tea Ceremony

Three people, a host, an elder, and a youth participate in the ceremony. There are four tasks they perform, listed in ascending order of importance;

Feeding the fire
Serving cakes
Serving tea
Reading poetry

The host performs all the tasks at the beginning of the ceremony, and the tasks are transferred back and forth among the participants until all the tasks are performed by the youth, at which time the ceremony is completed.

There are two constraints on the movement of tasks, only one task-the least important a person is performing-may be moved, and no person may receive a new task unless it is less important than any task they perform at the time.

The object of the Tea Ceremony game is to transfer the four tasks from the host to the youth in the fewest number of moves.

ToH Version 6: The Harbour Problem

There are three canals in the international harbour. Each canal has moorings for four ships. Each of these moorings is controlled by one of four countries;-

Britain

America

Russia

Denmark

Because of their patriotic tendencies no country will allow its moorings to be used by any other country.

There are four ships from four countries moored appropriately in the west canal. Unfortunately they all have to be moved to their appropriate moorings in the north canal in order that essential repairs can be carried out.

The problem is made worse because there is only one pilot in charge of all the harbour traffic so only one ship can be moved at a time. Furthermore, apart from the moorings in the north, west and east canal there are no other moorings because of strong tides. Can you show how the pilot can move the ships to the north canal?

ToH Version 7: The Art Gallery

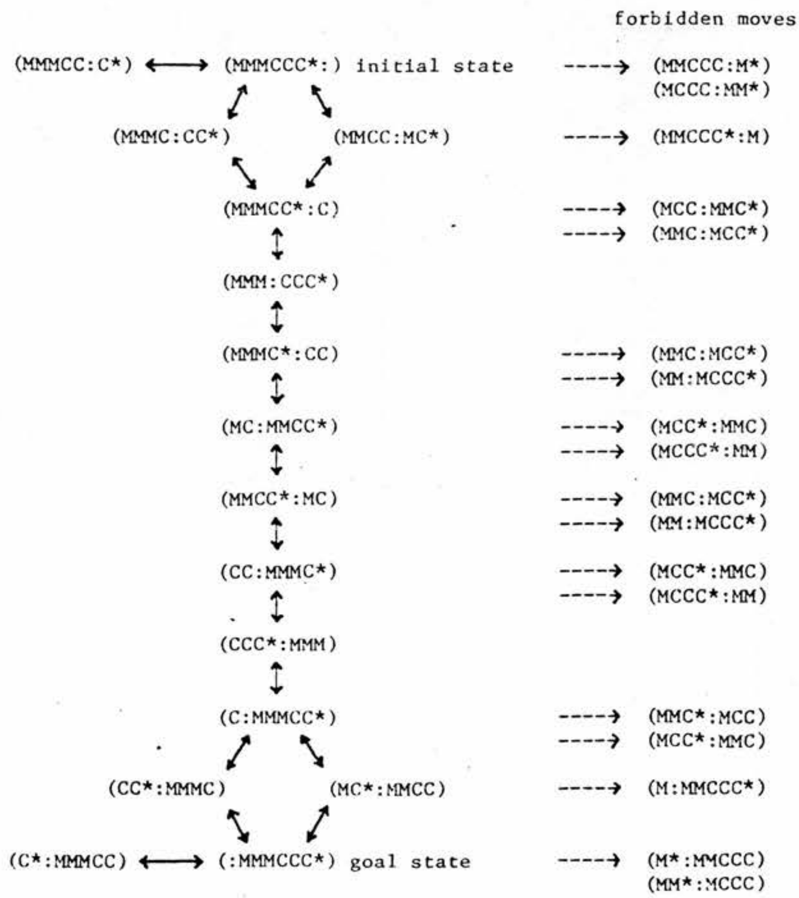
In the art gallery were three pedestals. The curator had placed a piece of modern sculpture on the lowest of the three pedestals. The sculpture consisted of four sections of different sizes all stacked together.

One day the sculptor arrived to view his masterpiece. On seeing his work on the lowest pedestal he complained, saying his work of art should be on the highest pedestal. He told the curator to move the sculpture, but warned him not to put any piece on the floor for fear of dirtying the fine colours, and not to rest any section on a smaller section in case the piece got damaged.

The curator was alone and as the sculpture was so heavy he had to move only one section at a time. And he could not rest two sections side by side because the pedestals were too small.

When the sculptor returned he found that the job had been properly done without dirtying or damaging the pieces. How did the curator carry out this complicated task?

State Space For The Missionaries And Cannibals Type Problems



MC Version 1: Missionaries And Cannibals

Three missionaries and three cannibals are on one bank of a river with a rowboat that will hold at most two people. How can they cross to the other side of the river, in such a manner that missionaries are never outnumbered by cannibals on either riverbank.

MC Version 2: The Jealous Husbands

Three jealous husbands and their wives, having to cross a river at a ferry, find that the boat is so small that it can contain no more than two people. Find the simplest schedule of crossing that will permit all six people to cross the river so none of the women shall be left in company with any of the men unless her husband is present. It is assumed that all passengers unboard before the next trip and at least one person has to be in the boat for each crossing.

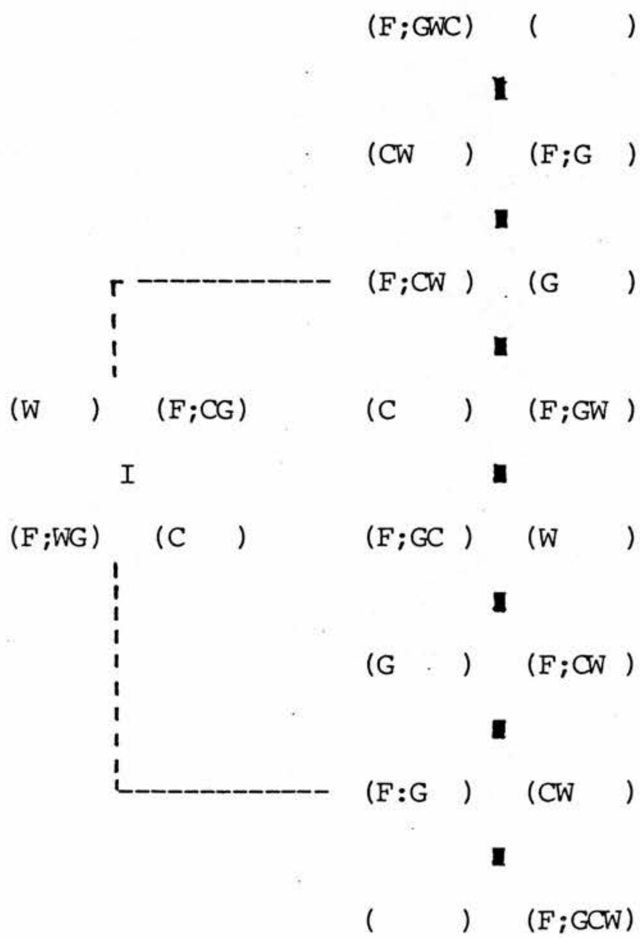
The State Space For The Goat, Wolf And Cabbage Problem

(Diagram Overleaf)

A farmer wishes to take his wolf, his cabbage, and his goat from the farmyard to market. He has use of a truck that is so small that it can only carry himself and one other item. So he must make more than one journey.

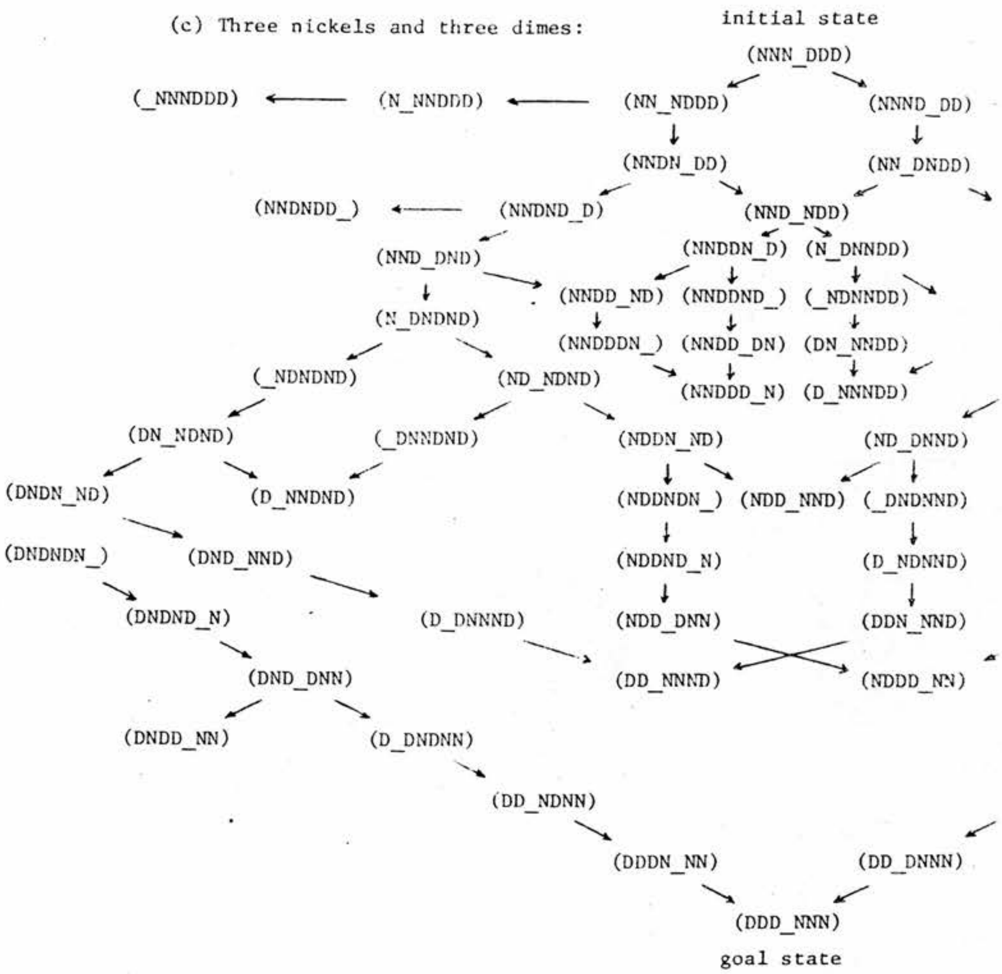
There is a problem however, if the goat is left alone with the wolf, the wolf will eat the goat. If the goat is left alone with the cabbage it will eat the cabbage. If the farmer is with the animals they will behave themselves.

Can you show how the farmer might get all his goods to market safely?



State Space for the the Goat/Wolf/Cabbage Problem

The State Space For The Three Nickels And Three Dimes

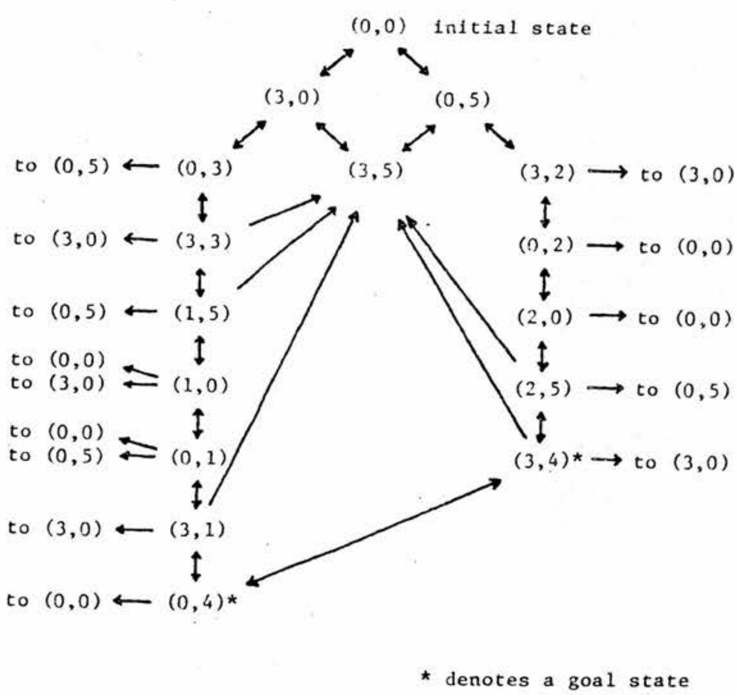


Version 1: The Shuntyard

The seven yellow circles in front of you labelled A-G are platforms in a shunting yard. The red lines are railway tracks. The three white trains at platforms A, B, and C, are required to move to platforms E, F, and G. The red lines are railway tracks. The three white trains at platforms A, B, and C, are required to move to platforms E, F, and G. Likewise the three black trains at platforms E, F, and G, are required to move to platforms A, B, and C.

Only one train is allowed at any platform at once. Trains may move in either direction along the track, but there is only one driver so only one train can move at once. Also it is against company regulations to leave trains anywhere except at platforms. If you were the driver how would you swap the trains over?

The State Space For The Two Pails Problem



Version1; The Sugar Refinaries

At the sugar refinary there is a large silo of sugar and two tankers. The tankers are of different sizes, the large one holds five tons of sugar when full, the smaller one three tons of sugar. Consequently it is easy to measure out three or five tons from the silo but other amounts require shrewd use of the silo and tankers at your disposal.

An order has come in for four tons of sugar, to be delivered to a factory in town in the large tanker. Your task is to find a means of loading exactly four tons of sugar from the silo into the large tanker. Can you show how this can be done with the silo and tankers at your disposal?

APPENDIX B

SOME ADDITIONAL EXPERIMENTS

Appendix B

Some additional experiments

Experiments with State Space Tasks

The experiments reported here were designed as follow-up studies to the first experiment described in the body of the Thesis (Chapter 3). In particular, the first experiment to be described is an attempt to replicate the failure to obtain any transfer across the isomorphic tasks in experiment 1 of the thesis. Also it includes a replication of the transfer that was finally obtained on these tasks using the MSP criteria. The second two experiments were designed to assess some of the factors assumed to affect the general (non-isomorphic) transfer obtained in experiment 1.

Experiment 1a

In this experiment two groups of subjects were required to solve the four ToH problems from experiment 1. In order to assess the possibility that the kind of instructions which the subjects were given might affect the functional criteria that they assumed to be at work in the experiment. The subjects were divided into two groups. One group were given instructions designed to focus their attention on the structural characterisitics of the problem, emphasising the requirement to solve the problem in as few a number of moves as possible. The second group were given the same instructions as subjects in experement 1 which merely advised them to solve the problems as quickly as was compatible with the avoidance of unnecessary errors.

Method

Task Materials

The four ToH problems used in experiment 1 of the thesis were also used in this study.

Subjects

Were 30 3rd-year psychology undergraduates took part in this experiment as part of their practical course. Subjects were of both sexes.

Design and Procedure

Subjects were divided into two groups, group 1 received the four ToH problems in a counterbalanced order as in experiment 1 of the thesis, and were instructed that they would be required to solve a few simple problems and that they should attempt to solve the problems as quickly as was consistent with not making needless errors. Thus the subjects of this group solved the four problems in similar conditions to those of experiment 1. Group 2 also solved the four problems in a counterbalanced order, but in this group received instructions designed to draw their attention to the structural characteristics of the problem by asking them to solve the problem in as few a number of moves as possible emphasising that time was not important.

Results and Discussion

Figures 1.la&b (opposite) present mean moves to solution and mean solution time for the two groups solving the four problems. s can be seen in terms of moves -to- solution group 2 show an overall better peformance than group 1 but an analysis of variance on these data show the effect of group just failing to reach significance ($p= 0.0578$), there was however a significant effect of practice ($p=0.012$), but by inspection it can be inferred that this may well be due to the very high score of group 1 on problem 2. The interaction between group and practice was not significant ($p > 0.05$).

Analyses of variance on the solution time data revealed that again there was no effect of group ($p = 0.7314$), but a significant effect of practice ($p = 0.0094$); and no interaction between group and practice ($p > 0.05$). Once again however it would seem likely that the high score for group 1 on problem 2 was responsible for the effects.

These results suggest in accordance with the experiments reported in the thesis, that simple exposure to structural isomorphism does not gaurantee that such invariances will be used by the subject to produce optimal solutions, even when subjects have been explicitly asked to attempt to solve the problems in as few a number of moves as possible. Any transfer effects that does occur in this procedure is singularly unimpressive. Certainly there is no indication of a shift towards minimum path solutions, even for the structure-oriented group. It appears that a more stringent procedure than simply asking the subjects will be required if such transfer is going to be obtained at all.

Experiment 1b

In this experiment the subjects of experiment 1a returned 1 week later. In this experiment however, both groups of subjects solved the same problems under the same experimental conditions, namely all subject solved three of the ToH problems they had solved the previous week, but this time the solved the problems under the MSP criteria described in experiment 2 of the thesis.

Method

Task Materials

Because of time constraints and the possibility that criterion procedures may take large amounts of time, it was decided that only 3 ToH problems would be used for this study. Thus 3 of the ToH problems solved by subjects in the previous week were used in this experment, The problem called ´computers´ being dropped (see Appendix A).

Subjects

Of the 30 subjects participating in the previous week´s experiment, 24 were able to attend the second week of testing.

Procedure

The 3 problems were presented to all subjects under the same conditions but in a counterbalanced order. All subjects received the ´solution time´ instructions from the previous week. The subjects were asked to solve each problem in turn, they were not explicitly informed of the criterion of success in operation but if subjects failed to solve a problem in the minimum number of moves possible (15) they were asked to solve the problem again, this time attempting to solve the problem in fewer moves. As in previous experiments moves -to- solution, and solution time, were recorded in addition to the number of attempts subjects needed to reach criterion.

Results

Table 2.1 (opposite) presents the mean number of attempts required to reach the MSP criterion, and figure 2.1 a&b (opposite) present the total moves and time to reach criterion.

As can be seen from this table even with the MSP criteria in force (and even after having previous experience with these problems), subjects were still requiring over 2 attempts to reach the criterion even on thie last problem. However there did appear to be some reduction not only in the number of attempts to reach solution but also in the time and moves to solution. Using non-parametric analyses of variance however, there was no significant effect on any of the three measures.

But in contrast to the experiment of the previous week, all subjects did achieve a minimum solution path, and the number of attempts required to do this was well within the number of attempts that subjects had made on the problems in week 1. That is to say that subjects in week 1 were essentially required to make 4 attempts at solving the ToH problem structure by solving each of 4 isomorphs once only. In this experiment subjects, even on their first problem required less than 3 attempts to reach criterion. From this comparison between the inter-problem transfer of week 1 and the intra-problem transfer of week 2, it is clear that the structure of the ToH problem is more readily asssimilated with deep exposure to a single problem than with shallow exposure to the class of isomorphic problems.

Thus it is clear that a high frequency of changes in the surface structure form of isomorphic problems with frequency of exposure to any one problem is less likely to produce a structurally mediated level of problem solving than a low frequency of exposure to surface structure variation within the context of a high frequency of repetition on a single form. While these results concur with the results in the area of ´concept formation´ studies, particularly those of Morrisett and Hovland for example (see Chapter 1).

<u>Experiment 2</u>

Given the apparent importance of criterion of mastery as to the type of transfer obtained on these tasks and the suggestion that expert performance on a single task might inhibit more general transfer (cf. experiments 1 and 2 of the thesis), the third experiment to be reported in this appendix assesses the effects of criterion procedures on transfer between two non-isomorphic state space tasks, namely the MC problem and the SR problem. In particular since solution time seemed to be the most important index of this general transfer, subjects in this experiment were run to a solution time criterion on one of the problems and then transferred to the second non-isomorphic problem.

Method

Task Materials

The two most difficult problems from experiment 1 were used in this study, namely the MC and the SR task.

Subjects

Were 8 undergraduates of both sexes attending first year psychology lectures in the Department of Psychology.

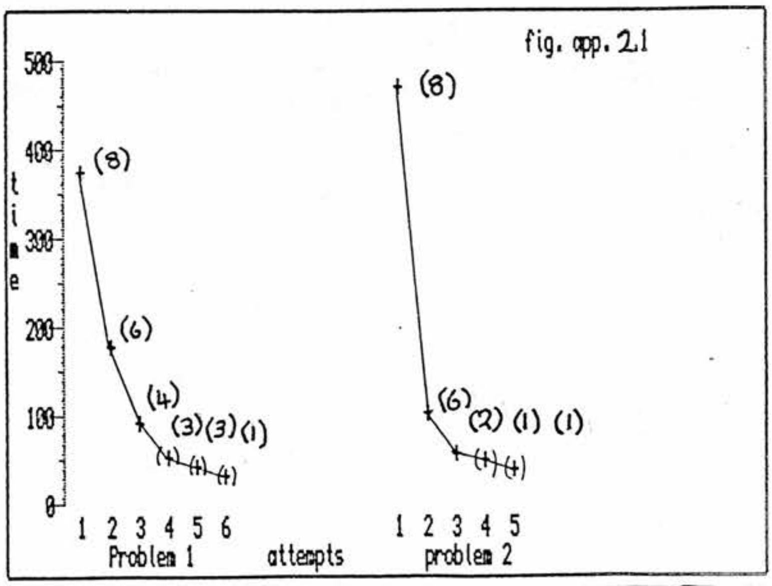
Design and Procedure

The mean solution times for the MC and SR problems after 7 antecedent prolems in experiment 1 of this thesis was 200 seconds. The subjects in this experiment were required to solve each problem to a criterion of 3 succesive solutions of less than 200 seconds. Subjects were encouraged to improve their solution times by "possibly reducing your thinking time or the number of moves you make". They were given feedback at the end of each attempt as to whether they had succeeded in reducing their solution time.

The subjects were divided into two groups one group solved the SR problem to criterion and then went on to the MC problem, the other group solved the MC first follwed by the SR problem. Both problems were solved to criterion.

Results and Discusion

The mean solution times for the repeated presentation of both problems for both groups is presented in figure 2.1.



Appendix figure 2.1. Reduction in solution time on each of two non-isomorphic problems solved to solution time criterion. Bracketed figures indicate number of subjects contributing to the mean.

Considering intra-problem transfer, T-tests on both problems for both groups reveal that the reduction in solution time is significant ($p < 0.05$). As can be seen by inspection however there is little indication of inter-problem transfer and the slight indication of negative transfer between the problem solved first and the problem solved second is not significant when the first attempt solution times on each problem are compared with t-tests ($p > 0.05$).

The results seem to suggest that deep exposure to a single problem reduces the possibility of transfer to other problems It must be noted however that this paradigm is a simple train and test paradigm and that with a multiple problem procedure under similar conditions would have been more illuminating in this respect.

Experiment 3

In experiment 1 of the thesis it will be recalled that it was argued there that the best predictors of transfer were not the structural relationships that existed between tasks modulo their state-space decomposition, but rather that the best predictor of transfer appeared to be the relative difficulty of the problems solved (determined by reference to the solution times of novice subjects solving these problems), and also the number of problems solved. This experiment assesses the effects of varying the number and difficulty of antecedent problems on subjects' performance on the most difficult Missionaries and cannibals problem.

Method

Task Materials

The 3 problems which proved most difficult for naive subjects to solve (as indexed by solution times recorded in experiment 1 of this thesis), namely MC, SR, and Rockets, and the two easiest problems, namely cakes and shuntyards were used in this experiment.

Subjects

18 subjects of both sexes were used in this experiment. The subjects were undergraduates attending 1st-year lectures in the department of psychology, University of Edinburgh. They were not paid for their services.

Design and Procedure

Subjects were divided into three groups, group 2e received the two easy problems followed by the MC problem, group 2h received the two hard problems followed by the the MC problem, and group 1h were subdivided, half received the SR problem followed by the MC problem, and half received the MC problem first followed by the SR problem. The performance of the naive subjects solving MC in this group was pooled with similar data from experiment 2 to provide a mean measure of naive performance on this task against which to compare the effects of different antecedent conditions.

From the above then, it can be seen that groups 2e and 2h recieved a set of 3 problems to solve in a prescribed order, while group 1h received a 2 problem set to solve in a prescribed order. But subjects did not simply solve 3 and 2 problems respectively, rather they solved their respective sets of problems repeatedly until a criterion performance of two succesive solutions of 200 seconds or less had been reached on the MC problem. This procedure allowed for an analysis of the effects of interpolating problems between repetitions of the MC problem. Subjects were given the same insructions as in experiment 2 and tested under similar conditions.

Results and Discussion

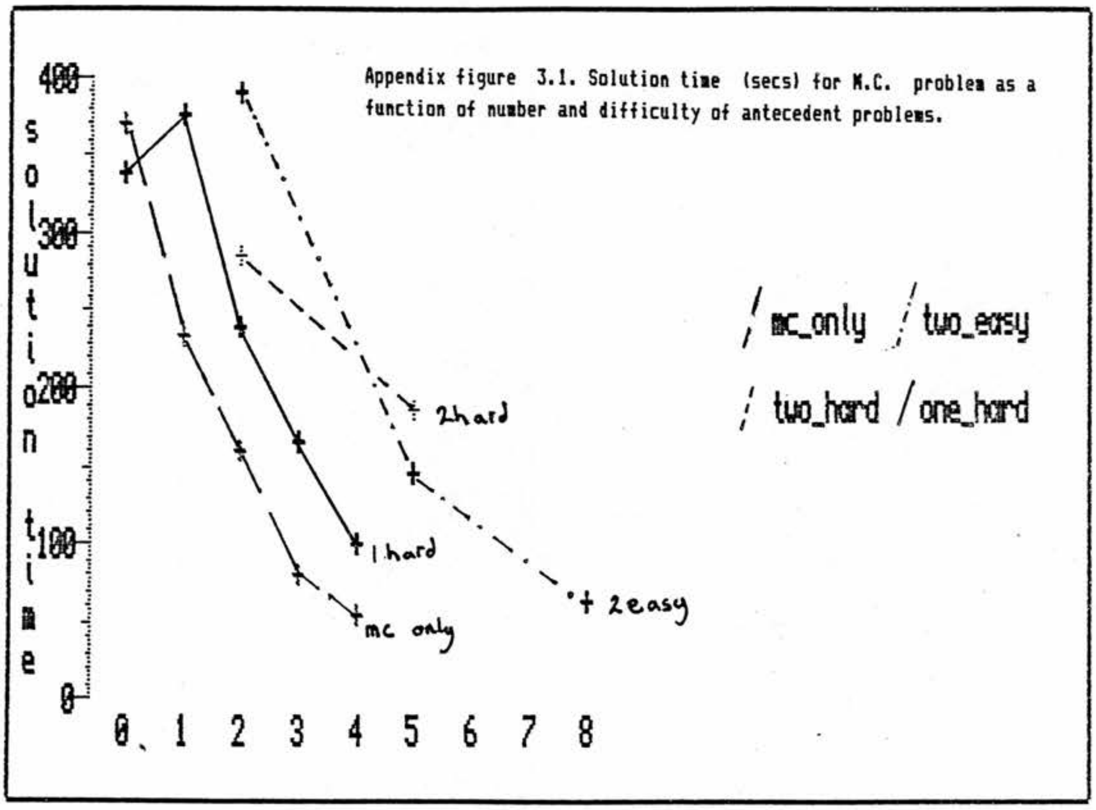
Table 3.1 presents meansolution times for subjects in each group on their first attempt only at solving the MC problem.

Mc Only	2 Easy antecedents	1 Hard antecedent	2 Hard antecedent
330	395	395	180

Appendix table 3.1. Solution time (secs) on the first attempt at the M.C. problem for the 4 experimental groups.

T-tests on these data suggest that only 2 hard antecedent problems significantly reduced the difficulty of the MC problem ($p < 0.05$), the solution times of groups 2e and 1h does not differ significantly from that of naive subjects ($p > 0.05$).

The effects of interpolating problems between successive solutions on the MC problem can be assessed by comparison to simple repetition to criterion on the MC task alone as was investigated in experiment 2. This comparison is presented in figure 3.1.



Appendix figure 3.1. Solution time (secs) for M.C. problem as a function of number and difficulty of antecedent problems.

The most striking affect is that the 2 hard non-isomorphic antecedent problems produce a level of performance on the MC problem comparable to that of 2 antecedent solutions of the the MC itself. In contrast 2 easy antecedents produce no such effect and the group with 1 hard antecedent problem falls neatly between these two. As can be seen the final performance level on all groups is comparable but in terms of the number of attempts to reach criterion, group 2h were best with 6 attempts to reach criterion, thus equalling the performance of subjects solving MC alone. Group

1h required 7 attempts while group 2e required 9 attempts.

These results suggest that the number and difficulty of antecedent problems is a good predictor of the extent of transfer to the MC task as suggested in experiment 1 of the thesis.

Experiment 4

A Further Experiment with Inference Problems

In the discussion of experiment 4 in the body of the thesis, it was argued firstly, that functional criteria are the determining factor in the subject's choice of strategy, and secondly that part of a subject's problem solving skill is the ability to determine what type of problem solving strategy is optimal given the functional demands of the problem solving situation. In this follow-up study evidence is provided that allows these points to be elaborated.

This experiment takes the 3 groups of subjects from experiment 4 into a new experimental condition in which they are required to solve the determinate problems from experiment 4; but this time with a partial question perspective. Thus for group 1 the problems have not changed but the functional perspective is different, for group 2 both the problems and the functional perspective change, and for group 3 only the problems change since in experiment 4 they solved indeterminate problems.

Thus in comparing these 3 groups under these different circumstances it will be possible to assess the extent to which subjects are sensitive to changes in the functional demands associated with these inference problems and secondly to assess the extent to which functional rather than structural criteria determine the generality of expertise developed in experiment 4.

Method

Task Materials

Subjects in this experiment solved the 32 determinate problems as used in experiment 4. In this experiment they were presented with the partial perspective questions as used for group 3 in experiment 4.

Subjects

37 of the 40 subjects that took part in experiment 4, returned 1 week later to act as subjects in this experiment.

Apparatus

As in experiment 4 problems were presented using an Apple II microcomputer which also recorded subjects´ premise reading times and question answering times.

Procedures

The procedures were identical to those of experiment 4 except this time all subjects made regular protocol reports throughout the session. 6 protocol reports were taken from each subject after the 1st, 4th,9th,17th,24th, and 32nd problems.

Table 4.1 presents the frequency of protocol reports falling into the categories of elimination and image formation reports. The subjects are group on the basis of the experimental condition that they had experienced in experiment 4 of the thesis.

	week 1 experience		
	Gp 1	Gp 2	Gp 3
Image formation	50%	50%	30%
Elimination	50%	50%	70%

Appendix table 4.1. Percentage of protocol reports for the two basic strategies as a function of week 1 experience.

As can be seen the group differences are small but there appears to be a higher frequency of elimination reports for the group that experienced the partial perspective in experiment 4. Of interest is that subjects who had previously experienced experimental conditions associated with the use of image strategies, while still maintaining a proportion of image reports, show a strong tendency towards the use of elimination. This suggests that even when subjects have a fully legitimate problem solving method readily available they will nevertheless an attempt to seek out the most optimal methods available. In this experiment 32% of subjects using an image strategy failed to show a shift towards the more optimal elimination method.

This suggests that the ability to recognise that existing methods are non-optimal and the ability to shift to new problem solving methods in the light of changes in the functional demands made on the subject may be a significant dimension of problem solving expertise along which people differ.

This analysis of the protocol reports suggests a number of things that might be expected to emerge in the analysis of premise integration time. Firstly given the higher incidence of elimination in group 3, then this group may show faster solution times that the other two groups as in experiment 4. However since the majority of subjects in the other blocks abandon the slower image method in favour of elimination it may be that although group 3 may have a faster solution time in the early blocks of problems, this advantage may disappear in latter blocks. Figure 4.1 (overleaf) shows the premise integration times of the three groups as a function of practice blocks.

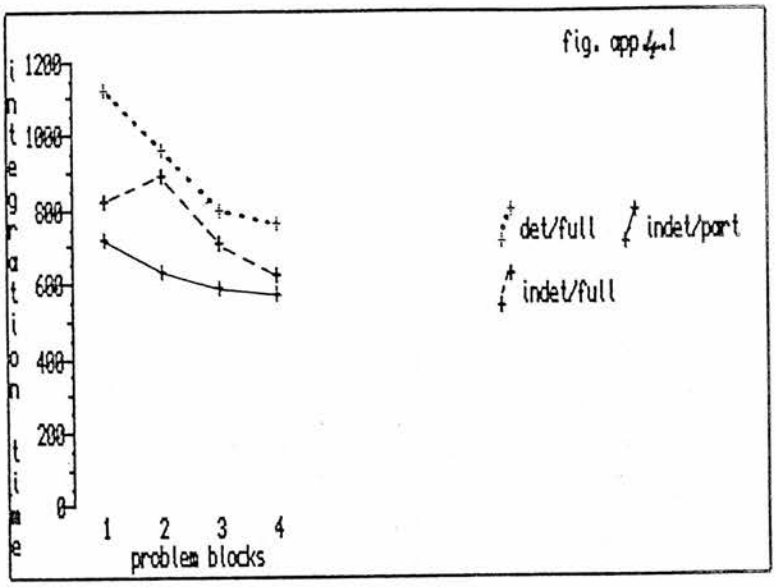


Figure 4.1. appendix. Premise integration time as a function of problem blocks for subjects grouped on the basis of week 1 experience.

By inspection it can be seen that, as anticipated, group 3 initially solve the problems much faster than the other two groups, particularly group 1. By comparing the times of these two groups on their first block of this experiment with that of their last block on experiment 4 (see figure 5.2.4) it is striking that both groups are performing at almost exactly the same levels across the two experiments. This is evidence for the transfer to this new experiment of previously successful problem solving methods. As can be seen from figure 4.1 however, all groups quickly converge on a very fast solution time as the number of problems solved increases. What is surprising is the relatively fast solution times for group 2 even on the initial block. It will be recalled that it this group that not only experienced a change in functional perspective between weeks 1 and 2 but also a change in the problems from indeterminate to determinate as well. It might have been expected then that they would have been slower than group 1. On the contrary they appear to perform consistently better than group 1. Thus it would appear that there is an asymmetry of transfer from determinate to indeterminate problem strucures.

An analysis of variance on these data reveal in fact that there is an effect of group just failing to reach significance ($F_{1,24} = 2.2$; $p = 0.0600$). This was so even when groups 1 and 3 were considered in isolation ($F_{1,24} = 2.6$; $p = 0.0526$), there was however a significant effect of practice blocks, ($F_{15,540} = 7.07$; $p < 0.0001$), and a significant group by practice interaction . This interaction supports the suggestion that the groups converge over practice blocks and if group differences on early blocks were to be considered in isolation, group effects may emerge.

Taken together the results of the premise integration time analysis and the protocol analysis suggest once again that functional rather than structural criteria are the major factors in determining the nature of the representational space that subjects adopt. Furthermore, even subjects that have a valid problem solving method based on previous experience with this type of problems will, nevertheless shift away from such methods when the functional perspective changes so as make then less than optimal. The effects of functional perspective appear to overide any advantages that might have been conferred upon the subject by virtue of prior experience to structurally identical problems. Finally the ability to track changes in the functional demands that the experiment places upon the subject and make appropriate adjustments to the problem solving method, appears to be a factor that might distinguish between good and poor problem solvers. A further analysis of solution times comparing static with more fluid problem solvers may prove illuminating in this respect.

APPENDIX C

SUMMARY TABLES FOR THE MAIN ANALYSES OF THE THESIS

Experiment 1: Anova 1.1 (cf table 3.1.2)

	source	S.S	D.F.	M.S.	F.	P.
	mean	5375510.63	1	5375510.63	257.46	0.0000
1	group	138666.95	1	138666.95	6.64	0.0219
	error	292305.30	14	20878.95		
	set	197899.13	1	198899.13	6.19	0.0261
2	SxG	307230.01	1	307230.01	9.61	0.0078
	error	447663.73	14	31975.98		
	practice	181205.90	3	60401.97	1.26	0.3005
3	PxG	6178.71	3	2059.57	0.04	0.9880
	error					
	SxP	4118.77	3	1372.92	0.03	0.9914
4	SxPxG	10715.27	3	3571.76	0.09	0.9659
	error	1693029.56	42	40310.23	0.09	0.9659

Experiment 1: Anova 1.2 (cf table 3.1.2)

	source	S.S	D.F.	M.S.	F.	P.
	mean	3818116.00	1	3818116.00	96.08	0.0000
1	group	429352.56	1	429352.56	10.80	0.0054
	error	556359.44	14	39739.96		
	practice	106272.13	3	35424.04	0.49	0.6914
2	PxG	3636.56	3	1212.18	0.02	0.9970
	error	3039005.31	42	72357.27		

Experiment 1: Anova 1.3 (cf. table 3.1.2)

	source	S.S	D.F.	M.S.	F.	P.
	mean	1755293.77	1	1755293.77	133.84	0.0000
1	group	16544.39	1	16544.39	1.26	0.2803
	error	183609.59	14	13114.97		
	practice	79052.55	3	26350.85	1.66	0.1908
2	PxG	13257.42	3	4419.14	0.28	0.8410
	error	667935.28	42	15903.22		

Experiment 1: Anova 1.4 (cf table 3.1.2)

source	S.S	D.F.	M.S.	F.	P.
mean	3620457.56	1	3620457.56	136.54	0.0000
error	185614.44	7	26516.35		
set	499142.25	1	499142.25	12.01	0.0105
error	290948.25	7	41564.03		
practice	80181.31	3	26727.10	0.35	0.7874
error	1589647.68	21	75697.51		
SxP	11041.88	3	3680.62	0.06	0..9813
error	1337422.63	21	63686.79		

Experiment 1: Anova 1.5 (cf table 3.1.2)

source	S.S	D.F.	M.S.	F.	P.
mean	1893720.02	1	1893720.02	124.25	0.0000
error	106690.86	7	15241.55		
set	5986.89	1	5986.89	0.27	0.6210
error	156715.48	7	22387.93		
practice	107203.30	3	35734.43	1.77	0.1839
error	424263.32	21	20203.01		
SxP	3792.17	3	1264.06	0.07	0.9730
error	355606.95	21	16933.66		

Experiment 1: Anova 1.6 (cf figure 3.1.2)

source	S.S	D.F.	M.S.	F.	P.
mean	2699449.00	1	2699449.00	156.70	0.0000
position	84535.56	1	84565.56	4.91	0.0686
error	103358.69	6	17226.45		
set	173056.00	1	173056.00	79.03	0.0001
SxP	66693.06	1	66693.06	30.46	0.0015
error	13138.19	6	2189.70		
problem	410388.12	3	136796..04	13.81	0.0001
Pr x Ps	67115.8125	3	22371.94	2.26	0.1164
error	178250.81	18	9902.82		
Pr xS	351468.12	3	117156.04	10.30	0.0004
Pr xSx Ps	124675.31	3	41558.44	3.65	0.0323
error	204785.31	18	11376.96		

Experiment 1: Anova 1.7 (cf figure 3.1.3)

source	S.S	D.F.	M.S.	F.	P.
mean	5151.13	1	5151.13	82.03	0.0000
group	2.53	1	2.53	0.04	0.8438
error	879.09	14	62.79		
practice	261.75	7	37.39	0.88	0.5221
PxG	620.34	7	88.62	2.10	0.0510
error	4145.16	98	42.30		

Experiment 1: Anova 1.8 (cf table 3.1.4)

source	S.S	D.F.	M.S.	F.	P.
mean	2475.06	1	2474.06	173.31	0.0000
position	49.00	1	49.00	3.43	0.1134
error	85.69	6	14.28		
problem	808.94	7	115.56	9.43	0.0000
POSxPROB	186.5	7	26.64	2.17	0.0000
error	514.81	42	12.26		

Experiment 1: Anova 1.9 (cf table 3.1.4)

source	S.S	D.F.	M.S.	F.	P.
mean	2295.03	1	2295.03	260.12	0.0000
position	26.28	1	126.28	2.98	0.1351
error	52.94	6	8.82		
problem	228.09	3	76.03	4.23	0.0198
POSxPROB	161.34	3	53.78	2.99	0.0581
error	323.31	18	17.96		

Experiment 1: Anova 1.10 (cf table 3.1.4)

source	S.S	D.F.	M.S.	F.	P.
mean	504.03	1	504.03	58.23	0.0003
position	22.78	1	22.78	2.63	0.1559
error	51.94	6	8.86		
problem	256.84	3	85.61	8.94	0.0008
POSxPROB	25.09	3	8.36	0.87	0.4730
error	172.31	18	9.57		

Experiment 2: Anova 2.1 (cf figure 3.2.1)

source	S.S	D.F.	M.S.	F.	P.
mean	5077481.74	1	5077481.74001	74.21	0.0000
group	20584.35	1	20584.35	0.30	0.5891
error	1436739.54	21	68416.17		
practice	987137.98	5	195627.59	8.44	0.0000
PxG	293883.54	5	58776.71	2.54	0.0329
error	2433228.55	105	23173.61		

Experiment 2: Anova 2.2 (cf figure 3.2.1)

source	S.S	D.F.	M.S.	F.	P.
mean	3002883.56	1	3002883.56	435.54	0.0001
error	929462.78	11	84496.62		
practice	1198119.06	5	239623.92	6.3	0.0001
error	2091882.06	55	38034.22		

Experiment 2: Anova 2.3 (cf figure 3.2.1)

source	S.S	D.F.	M.S.	F.	P.
mean	2133003.41	1	2133003.41	42.05	0.0000
error	507276.76	10	50727.68		
practice	120744.32	5	24148.86	3.54	0.0081
error	341346.51	50	6826.93		

Experiment 2: Anova 2.4 (cf figure 3.2.2)

source	S.S	D.F.	M.S.	F.	P.
mean	1693514.40	1	1693514.40	69.92	0.0000
group	11524.55	1	11524.55	0.48	0.4979
error	508638.76	21	24220.89		
practice	16827.48	5	33365.49	4.41	0.0011
PxG	20120.47	5	4024.09	0.53	0.7513
error	793639.79	105	75588.47		

Experiment 2: Anova 2.5 (cf figure 3.2.3)

source	S.S	D.F.	M.S.	F.	P.
mean	135981.13	1	135981.13	42.76	0.0000
error	34979.04	11	3179.93		
practice	24480.29	5	4896.06	5.80	0.0002
error	46422.54	55	844.04		

Experiment 2: Anova 2.6 (cf figure 3.2.3)

source	S.S	D.F.	M.S.	F.	P.
mean	134371.00	1	134371.00	371.43	0.0000
error	3617.70	10	361.77		
practice	864.48	5	172.90	0.64	0.6575
error	13154.85	50	263.09		

Experiment 2: Anova 2.7 (cf table 3.2.3)

source	S.S	D.F.	M.S.	F.	P.
mean	7436.13	1	7436.13	191.10	0.0000
group	1.04	1	1.04	0.03	0.8720
error	778.26	20	38.92		
problem	1322.00	1	1322.00	61.54	0.0000
PxG	1.82	1	1.82	0.08	0.7741
error	429.66	20	21.48		

Experiment 2: Anova 2.8 (cf table 3.2.3)

source	S.S	D.F.	M.S.	F.	P.
mean	4621981.73	1	4621981.73	83.85	0.0000
group	1204.64	1	1204.64	0.02	0.8840
error	1102431.66	20	55121.58		
problem	82270.64	1	82270.64	1.89	0.1849
PxG	7512.27	1	7512.27	0.17	0.6826
error	872459.66	20	43622.98		

Experiment 3: Anova 3.1 (cf table 6.3 and figure 6.1)

source	S.S	D.F.	M.S.	F.	P.
mean	14326460.86	1	14326460.86	214.70	0.0000
group	303976.70	1	303976.70	4.56	0.0417
error	1868409.00	28	66728.89		
set	239578.74	1	239578.74	18.98	0.0002
SxG	1252019.99	1	1252019.99	99.20	0.0000
error	353397.68	28	12621.35		
practice	1096317.31	15	73087.82	7.59	0.0000
PxG	80018.15	15	5334.54	0.55	0.9086
error	4043162.61	420	9626.58		
SxP	135472.33	15	9031.49	1.10	0.3537
SxPxG	254957.88	15	16997.19	2.07	0.0105
error	3447684.49	420	8208.77		

Experiment 3: Anova 3.2 (cf figure 6.1)

source	S.S	D.F.	M.S.	F.	P.
mean	8244485.16	1	8244485.16	139.21	0.0000
error	769904.71	13	59223.44		
practice	917769.70	15	61184.64	3.48	0.0000
error	3429176.43	195	17585.52		

Experiment 3: Anova 3.3 (cf figure 5.1.1)

source	S.S	D.F.	M.S.	F.	P.
mean	1816598.53	1	1816598.54	85.68	0.0000
error	318014.53	15	21200.97		
practice	229450.15	15	15296.68	2.58	0.0014
error	1336583.78	225	5940.37		

Experiment 3: Anova 3.4 (cf figure 5.1.1)

source	S.S	D.F.	M.S.	F.	P.
mean	39975.00	1	3997500.39	82.53	0.0000
error	726560.36	15	48437.36		
practice	277757.86	15	18517.19	2.28	0.0052
error	1829715.39	225	8132.06		

Experiment 4: Anova 4.1 (cf figure 5.2.2)

source	S.S	D.F.	M.S.	F.	P.
mean	18256886.25	1	18256886.25	506.14	0.0000
group	26810.52	1	26810.52	0.74	0.3975
error	829632.66	23	36070.98		
practice	319173.06	3	106391.02	15.95	0.0000
PxG	40212.63	3	13404.21	2.01	0.1207
error	460389.52	69	6672.312		
IDE	156276.01	3	52092.00	12.29	0.0000
IxG	2367.01	3	789.00	0.19	0.9055
error	292569.66	69	4240.14		
PxI	73989.97	9	8221.11	1.65	0.1019
PxIxG	34662.00	9	3851.33	0.78	0.6396
error	1028606.62	207	4969.11		

Experiment 4: Anova 4.2 (cf figure 5.2.3)

source	S.S	D.F.	M.S.	F.	P.
mean	390671936.16	1	390671936.16	307.61	0.0000
group	9187542.88	2	4593771.44	3.62	0.0373
error	4451386.78	35	1270039.62		
practice	4906296.79	3	1635432.26	16.76	0.0000
PxG	1102836.46	6	183806.08	1.88	0.0904
error	10246094.57	105	97581.85		
figure	1364678.19	3	454892.73	10.35	0.0000
FxG	1065046.00	6	177507.67	4.04	0.0011
error	461497.137	105	43952.16		
PxF	871271.95	9	96807.99	2.19	0.0228
PxFxG	911634.18	18	50646.34	1.14	0.3087
error	13952079.79	315	44292.32		

Experiment 4: Anova 4.3 (cf figure 5.2.3)

source	S.S	D.F.	M.S.	F.	P.
mean	328334680.43	1	328334680.43	242.28	0.0000
group	90868.89	1	90868.89	0.07	0.7979
error	32523854.00	24	1355160.58		
practice	1902119.35	3	634039.78	7.59	0.0002
PxG	98957.67	3	32985.89	0.39	0.7572
error	60167881.69	72	83566.41		
figure	1577419.55	3	525806.51	12.93	0.0000
FxG	263117.55	3	87705.85	2.16	0.1007
error	2928896.17	72	40679.11		
PxF	481066.00	9	53451.77	1.37	0.2555
PxFxG	395781.67	9	43975.74	1.13	0.3448
error	8429177.95	216	39123.97		

Experiment 4: Anova 4.4 (cf figure 5.2.4)

source	S.S	D.F.	M.S.	F.	P.
mean	75973085.44	1	75973085.44	70.07	0.0000
practice	3876438.78	3	1292146.25	10.08	0.0001
error	4229312.87	33	128160.99		
figure	595052.94	3	198358.98	3.88	0.0176
PxF	896310.87	9	99590.09	1.79	0.0804
error	5522901.84	99	55786.89		

Experiment 4: Anova 4.5 (cf figure 5.2.6)

source	S.S	D.F.	M.S.	F.	P.
mean	95398383.81	1	95398383.81	377.41	0.0000
group	2587651.87	1	3587651.87	5.12	0.0111
error	9099745.24	36	2552770.70		
proactice	875501.12	3	281833.71	19.55	0.0000
PxG	248413.94	6	41402.32	2.77	0.0151
error	1612372.12	69108	14929.37		
figure	180467.49	3	60155.83	6.97	0.0002
FxG	161568.58	6	26928.10	3.12	0.0074
error	931856.72	108	8628.30		
PxF	95120.71	9	10568.97	1.46	0.1620
PxFxG	137053.42	18	7614.08	1.085	0.4016
error	2346521.87	324	7242.35		
Premise	104554.32	3	34851.44	1.74	0.1635
GxPrem	305111.32	6	50851.89	2.54	0.0246
error	2165680.48	108	20052.60		
PxPrem	164698.18	9	18299.80	3.64	0.0002
GxPxPrem	98143.26	18	5452.40	1.09	0.3652
error	1627112.15	324	5021.95		
FxPrem	124792.94	9	13865.88	2.98	0.0020
FxGxPrem	127272.97	18	7070.72	1.52	0.0803
error	1505748.25	324	4647.37		
PxFxPrem	187034.71	27	6927.21	1.59	0.0288
PxFxGxPrem	264106.32	54	4890.86	1.12	0.2550
error	4230264.55	972	4352.12		

Experiment 4: Anova 4.6 (cf figure 5.2.6)

source	S.S	D.F.	M.S.	F.	P.
mean	82641887.76	1	82641887.76	319.31	0.0000
group	51992.30	1	51992.30	0.20	0.6579
error	6470372.45	25	258814.90		
practice	375282.43	3	125094.14	10.39	0.0000
PxG	86125.88	6	28708.63	2.39	0.0758
error	902602.09	75	12034.69		
figure	269627.20	3	89875.73	8.96	0.0000
FxG	8018.31	3	2672.77	0.27	0.8493
error	751954.03	75	10026.05		
PxF	88829.82	9	9869.98	1.24	0.2701
PxFxG	86606.22	9	9622.91	1.21	0.2869
error	1787293.22	225	7943.52		
Premise	77862.48	3	25954.16	1.15	0.3343
GxPrem	199968.20	3	66656.07	2.96	0.0378
error	2165680.48	108	20052.60		
PxPrem	73902	9	8211.34	1.39	0.1925
GxPxPrem	52914.11	9	5879.35	1.00	0.4433
error	1326742.36	225	5896.63		
FxPrem	147543.55	9	16393.72	3.15	0.0013
FxGxPrem	56702.78	9	6300.31	1.21	0.2886
error	1169681.56	225	5198.58		
PxFxPrem	154127.72	27	5708.43	1.2	0.2278
PxFxGxPrem	145299.73	27	5381.47	1.13	0.3001
error	3222342..93	675	4773.84		

Experiment 4: Anova 4.7 (cf figure 5.2.6)

source	S.S	D.F.	M.S.	F.	P.
mean	17367423.33	1	17367423.33	72.31	0.0000
practice	612033.10	3	204011.03	9.18	0.0001
error	733768.68	33	22235.41		
figure	83968.45	3	27989.48	5.00	0.0057
PxF	58340.07	9	6482.23	1.10	0.3702
error	583589.42	99	5894.44		
Premise	122791.03	3	40930.34	2.87	0.0513
PxPrem	1351669.74	9	15018.86	4.46	0.0001
FxPrem	58268.69	9	6474.298	1.79	0.0800
PxFxPrem	164051.19	27	6075.97	1.67	0.0223
error	1081559.35	297	3641.613		

Experiment 4: Anova 4.8 (cf figure 5.2.6)

source	S.S	D.F.	M.S.	F.	P.
mean	5604187.53	1	5604187.53	52.09	0.0000
practice	2802.20	1	2802.20	0.33	0.5752
error	92385.22	11	8398.66		
figure	25040.27	3	8346.76	2.36	0.0889
PxF	1768.38	3	589.46	0.17	0.9126
error	111245.89	33	3371.09		
Premise	11841.22	3	3947.07	0.58	0.6351
PxPrem	14804.35	3	4934.78	2.03	0.1282
FxPrem	28229.66	9	3136.63	1.14	0.3395
PxFxPrem	21879.76	9	2431.08	0.89	0.5339
error	269312.16	99	2720.324		

APPENDIX D

CLASSIFICATION OF SOME EXAMPLE PROTOCOLS

APPENDIX D

A Classification and Detailed Analysis of Some Example Protocols

This appendix presents example protocol reports from three of the experiments in the thesis. These reports have not been chosen at random, they have been chosen to highlight the classification processes outlined in the body of the thesis. The reports are a cross-section, some of which fall easily into the classificatory schema and some which do not fit so well. The protocol reports are presented in the order which reflects the importance that was placed on them in the body of the thesis. Thus protocols from experiment 4 are presented first, followed by those from experiment 3, and finally those from experiment 2 with the ToH problems.

For each of the experiments, a short summary of the procedures is presented, followed by an account of the classificatory schema for that experiment. Each of the categories into which the protocols are classified is described in some detail. This is followed by the protocol reports themselves. The protocols are presented on the left of the page while on the right, the category into which the report is assigned is noted down. Where necessary, the protocols are underlined to indicate which part of the report is responsible for its assignment to the category in question.

EXPERIMENT 4: Representation, Strategy and Expertise

Summary:

In this experiment, three groups of subjects solve thirty-two 5-term inference problems. In group 1, the problems were fully determinate (eg. A>B, B>C, C>D, D>E) and for each problem subjects were required to answer two comparative questions (eg. is A>B?). In group 2, subjects received the same questions but their problems were indeterminate (eg. A>B, A>C, C>D, D>E). In group 3, subjects received the same indeterminate premises as group 2, but instead of receiving two comparative questions after the premises, they received a single superlative (ie. Who is tallest?) before they received the premises.

Subjects provided a varying number of protocol reports throughout the session. For the reports presented below, the number corresponds to the problem immediately before the report was taken. In this experiment, whenever a subject reported using some form of mental image, he was explicitly asked to say whether this was a dimensionalised image, and if so, he was asked along which axis and in what direction the items were arranged. Subjects in the indeterminate groups were asked whether they noticed any ambiguity in the premises (cf. Ch. 5, pp. 153-154 for details).

The categories:

This experiment was concerned with a very constrained set of questions concerning the protocols, and the answers to these questions were self-evident in the protocols. The protocols are categorised according to whether some form of ELIMINATION or some form of IMAGE strategy was used.

Subjects reporting the use of an image are classified as to whether the image was GLOBAL or DIMENSIONALISED. ´Global´ refers to reports which simply mention "making a mental picture" without specifying whether this was some concrete image of the premises themselves, or the people described by the premises, rather than a spatial array of the names. In contrast, ´dimensionalised´ refers to reports which specify explicitly, that the subject was ordering the items of the premises along some dimension and in a particular direction, such as "Forming a mental picture of the tallest to smallest, from left to right". Reports of using a ´dimensionalised image´ fall into 4 sub-categories (L/R, R/L, T/B, B/T), according to whether they report ordering the items (starting with the tallest) from left to right or vice versa or from top to bottom or vice versa. There was no difficulty in determining which of these various categories the protocols fell into.

ELIMINATION, refers to reports in which the subject uses his knowledge of the question (group 3 only) to guide his search for relevant premise information. Because of the particular problem forms used in this experiment, a subject using this strategy is able to discover his answer after reading only two of the four premises. Subjects using this strategy do not report retaining all of the items in any form of image, rather they report remembering only "Who is tallest" and pay "only slight attention to other available info." Because of the great detail in subjects' reports of this strategy, there is no problem of classification. It must be pointed out that the elimination reported in this experiment is NOT the same as that reported in experiment 3, the distinctions between the two forms is made clear in the discussion of that experiment.

In addition to the main distinction between image and elimination, two other minor categories have been used, these are REHEARSAL AND MNEMONICS. 'Rehearsal' refers to reports often occurring in conjunction with image reports (cf. discussion of experiment 3), in which subjects report making an image and then verbally rehearsing the items. 'Mnemonics' refers to subjects who, instead of remembering the names in full, instead just remember the initial letters.

Finally, for groups 2 and 3, it is noted whether they recognise that the information with which they are presented is ambiguous.

Example Protocols:

REPORT	CATEGORY
SUBJ: N.M. (GP. 1 det/comp)	
1: Put them in order using the whole name Tall -> SHORT, LEFT -> RIGHT slotting names in order.	DIMEN. IMAGE: L/R
16: Rehearsal of names.	REHEARSAL
17: Task becoming easier - although no experimenting with a new method- stick to the same mental image of names L -> R,Tallest ->S.	DIMEN. IMAGE: L/R
32: By the last [question] task much easier faster presentation and less verbal recall providing allowed to give response immediately.	---
SUBJ: F.M. (GP. 1 det/comp)	
1: Forming a mental picture of the tallest to smallest from left to right.	DIMEN. IMAGE: L/R
16: Same method as before.	DIMEN. IMAGE: L/R
17: Same method as before.	DIMEN. IMAGE: L/R
32: Same.	DIMEN. IMAGE: L/R
SUBJ: N.W. (GP. 2 indet/comp)	
1: Can´t remember order	--
9: Full names, kept tallest at the top shortest at the bottom	DIMEN. IMAGE: T/B
16: Ambiguous: Emma TALLEST Kate Helen Ruth June SHORTEST Remembering names	RECOG. AMBIGUITY DIMEN. IMAGE: T/B

17:	A E M G J Rehearse frequently in the same order. TALLEST	DIMEN. IMAGE: T/B + REHEARSAL
25:	Helen TALLEST Pat Jane Mary Gill Rehearse names frequently. Concerned as to what the experiment is really for.	DIMEN. IMAGE: T/B + REHEARSAL
32:	T H G B AMBIGUOUS P Still remembering the same way.	DIMEN. IMAGE: T/B +RECOG. AMBIGUITY + REHEARSAL

SUBJ: C.C. (GP. 2 indet/comp.)

1:	I tried to memorize and picture the situations presented.	GLOBAL IMAGE
4:	Same processes used; also a bit of lucky guessing.	--
8:	The shorter one in my mind is at the left while the taller one is at the right. When asked the question which one is shorter or taller, I automatically see the taller one first.	DIMEN. IMAGE: L/R
9:	TALLER -> SHORTER Sue Kate Mary Jane I´m now picturing people I know and am now associating names with these people.	DIMEN. IMAGE: L/R +MNEMONIC
12:	I´m repeating each statement presented and comparing the two girls side by side, and also attempting to compare the last statement to the one	REHEARSAL +GLOBAL IMAGE

presented. At the end I try to picture them as a group in order to make comparisions more easily.

16: Because most of the names were unfamiliar this time (ie. couldn´t associate them with friends), I used more memorization than before.	—
17: I´m still using the method of comparisons of names with faces I know. I am also using repeating + memorization of those names I can´t picture (ie. I´ve no friends of the name IAN so I tried to memorise it.	REHEARSAL +GLOBAL IMAGE
20: No difference.	GLOBAL IMAGE + REHEARSAL
24: A lot easier for me to picture and remember the size order when put "so and so is TALLER THAN"	GLOBAL IMAGE + REHEARSAL
25: No difference though (I´m concentrating a bit less)	GLOBAL IMAGE + REHEARSAL
28: No difference (though the names made it easier in this one)	GLOBAL IMAGE + REHEARSAL
32: No difference.	GLOBAL IMAGE + REHEARSAL

SUBJ: S.O. (GP. 3 Indet/sup)

1: Remember who is tallest. Only pay slight attention to the other available info.	ELIMINATION
16: Same. But wait to make sure other statements are same type.	ELIMINATION
17: Can work out the answer from the first statement therefore ignore the others.	ELIMINATION
32: As before wait till after 2nd statement to make sure sequence is as expected.	ELIMINATION

SUBJ: J.K.

1: Emma, Helen, Kate, Pat, Ruth If question is "Who is tallest" + 1st statement is Emma is taller than.. then I remember Emma as being the tallest so far & I go on to the next statement. If no one else is taller than Emma then she´s the tallest. — ELIMINATION

16: The first two statement always involves the same name i.e. Emma is in both. I decide who is tallest or shortest from the 1st 2 statements and if that name is not mentioned again in the 3rd and 4th then he or she is the tallest/shortest. — ELIMINATION

17: Use strategy as above. — ELIMINATION

32: No ambiguity. Purely tried to answer question so ignored names other than the answer. — ELIMINATION

EXPERIMENT THREE: Three and Five- Term Transfer

Summary:

In this experiment two groups of subjects solved 32 inference problems. Subjects in group 1 solved 16 three-term series problems such as:

Alan is taller than Bob

Carl is shorter than Bob

Who is the tallest?

These were followed by 16 five-term series problems such as:

Alan is to the left of Bob

Carl is to the right of Bob

Edward is above David

Carl is to the left of David

Who is leftmost?

In group 2, subjects solved the 16 five-term problems first, followed by the 16 three-term problems. The problems were presented by an APPLE II microcomputer. Premises were presented simultaneously for 12 seconds in the case of three-term problems and 24 seconds in the case of five- term problems. Premises and question were presented separately, the computer logged the time taken for subjects to answer the question. Subjects were told to answer the problems as quickly as was consistent with a high degree of accuracy.

All subjects in both groups were required to provide 12 written protocol reports of how they solved the problem. These were taken after they had provided the answer to their, 1st, 4th, 8th, 9th, 12th, and 16th problems of each problem type, and before they began the next problem.

The categories:

As with experiment 4 above, the protocol reports were analysed according to 6 categories. (see chapter 5, pp. 127-134 and figures 5.1.2 & 5.1.3). The first two of these - **GLOBAL IMAGE** and **DIMENSIONALISED IMAGE** refer to subjects´ use of some form of spatial image device. Once again, this is indicated in the protocols, by specific reference to pictures or images or by reference to the placement of items in space. The two categories are distinguished by the fact that the second includes reports of "ordering" the names along a spatial dimension such that the ordering of the items along that dimension is in some way analogous the relational properties described by the pemises; For example the "topmost" or "leftmost" place in the spatial image, is reserved for the "tallest" or "smallest" person and so on.

By and large, it is a straightforward matter to decide whether these image strategies are being employed, but one particular difficulty worth note are the reports of ´list formation´; Here subjects report rehearsing a list of the items in order of height starting with the tallest (subj:21, below for example). It would seem that such a list has the characteristics of a dimensionalised image in as much as the ordering of the list is an analogue of the relational properties described in the premises, but it appears to be a verbal rather than imaginal strategy. Pending further enquiry, this list-formation strategy is categorised under **VERBAL REHEARSAL** rather than dimensionalised image. Finally we must note that with a small number of reports, while it is clear that the subject is using some form of image device, it is not clear whether it is a dimensionalised image. (see subj:32, report number 1; for example).

In contrast, category three- **REHEARSAL** is a straightforward category to score, since subjects explicitly use the term "rehearsal" or "verbal rehearsal" or "verbally recite".

The fourth category - **SCANNING** refers to a strategy a few subjects adopt in which they do not simply read the premises in the order presented. Instead, they read the first premise and then search the other premises for information which can be integrated directly with that obtained from the first premise. Thus the reading order for the problem A>B, D>E, B>C; would be 1,3,2. This usually occurs as part of the strategy for forming a dimensionalised image as below:

"start with 2 names located in space then look for others to put in the space directly before or after them." Because three-term problems only have two premises, this category of report is confined to subjects solving the five-term problems.

The fifth category- **ELIMINATION**, is to be distinguished from the elimination reported in experiment 4. Unlike experiment 4, subjects in this experiment do not use their knowledge of the question they are asked to guide their search through the premises for the correct answer- how could they since they are asked the question after reading the premises? Rather, they use a modified image device to remember only the end-anchor items. This is a valid strategy in this particular study because, although subjects do not know what question they will be asked, they do recognise that they are only being asked superlative questions and thus are only being interrogated about end-anchor items. While somewhat rare among our subjects elimination is an easy category to spot.

The final category- **others,** is a bin category for any other strategies that subjects use systematically. By and large these are mnemonic strategies such as remembering only the initial letters of the names and so on. One sub-category; **MNEMONIC; VISUAL,** refers to reports of scanning the premises back and forth. But unlike the **SCANNING** category above, subjects are not searching for a particular item, rather they appear to be using a visual rehearsal procedure which somehow helps them remember certain premise information.

Example Protocols:

REPORT	CATEGORY
SUBJ: 22	
5-term	
1: Guess	--
4: Picturing a list in order	DIMEN. IMAGE
8: Picturing a list of names in order of height	DIMEN. IMAGE
9: Repeating verbally an ordered list	REHEARSAL
12: Repeating a verbal list of the names in either ascending or descending order depending on the layout of the information given	REHEARSAL
16: Construct a list of names ordering each name following the instructions (verbal).	REHEARSAL
3-term	
17: Picture the names in space	GLOBAL IMAGE
20: Spatially locate places for names verbally repeat list of names placing them in spatial position in a certain order.	DIMEN. IMAGE +REHEARSAL
24: Start with 2 names located in space then look for others to put in the space just before or after them, then repeat the list and look for another in the list then I add the names which come above or below the list	DIMEN. IMAGE +SCANNING +REHEARSAL
25: Same	DIMEN. IMAGE +SCANNING +REHEARSAL
28: Same : Visual placing, verbal repetition of names.	DIMEN. IMAGE +SCANNING +REHEARSAL

32: Same: scan all sentences first then spatially locate + verbally rehearse the list.	DIMEN. IMAGE +SCANNING +REHEARSAL

Subj: 21

3-term

1: Sorted out order of height and then recited this order in to myself repeatedly	REHEARSAL
4: Same as before Sorted out order of height recited tallest 1st then 2nd tallest 2nd and so on in to myself repeatedly.	REHEARSAL
8: Same strategy used as before recite tallest 1st in to myself.	REHEARSAL
9: Recite tallest name 1st followed by 2nd tallest + then 3rd tallest in to myself	REHEARSAL
12: As before read through sentences and establish who is tallest then recite 2nd tallest + then 3rd in to myself	REHEARSAL
16: Same as before - no difference	REHEARSAL

5-Term

17: Total confusion- I´m hopeless with left and right!!!!! However, seemed to remember Barry was above someone else so plumped for him	--
20: Work out order of names from left to right + try to quickly note other name and whether above or below. Main strategy to remember L->R name sequence.	REHEARSAL
24: Same as before but as reciting look at statement involving above or below.	REHEARSAL +OTHERS: VISUAL
25: Same as before - no change	REHEARSAL +OTHERS: VISUAL

28: Same as before

REHEARSAL
OTHERS: VISUAL

32: Verbal rehearsal strategy scan sentences back and forth at end whilst reciting devote most attention to sentence involving above or below.

REHEARSAL
OTHERS: VISUAL

SUBJ: 52

5-term

1: Place people´s name in space according to relationship described read once, place and rehearse the arrangement.

DIMEN. IMAGE
+REHEARSAL

4: No change except a little more rehearsal of names (since so many being thrown out at me!) If the same names were being used I would find it easier.

DIMEN. IMAGE
+REHEARSAL

8: No change yet, but considering only remembering the outside names so less names to remember. This is because I now know the limited no. of questions being used!

DIMEN. IMAGE
+REHEARSAL
+ELIMINATION

9: Now remembering top, bottom, right and leftmost, I hope the questions don´t change!

ELIMINATION

12: This methods easier to remember names with, I think I´m getting faster.

--

16: Originally I was reading the lines from top to bottom fitting into space, somewhere in the first few I changed to take first name, attach second to it, scan for any other occurrence of first name etc. ie. taking one name and attaching others to it rather than having say 2 pairs of names and later joining them up.

DIMEN. IMAGE
+SCANNING

3-term

17: By visual representation of height as steps tallest being top step etc — DIMEN. IMAGE

```
 ----
  ----
     ----
```

With only 3 names don´t need to rehearse

20: No change. With lots of waving of arms (perhaps its the sign language I´ve been learning recently) — DIMEN. IMAGE

24: No change. With having to tag names onto the left, need more rehearsal of names can´t just add onto the end of the list — DIMEN. IMAGE +REHEARSAL

25: No change at all — DIMEN. IMAGE +REHEARSAL

28: Since rehearsal time quite long, started playing games, eg. Ann, Clare, Emma initials = ACE but I still remembered Ann Clare Emma. — OTHERS: MNEMONIC

32: Getting more visual again, rather than envisaging a list, envisaged statements separately considering them both after the question. — --

SUBJ: 32

5-term

1: Picture of relationships — GLOBAL IMAGE

4: Try to construct an image of names in a line — DIMEN. IMAGE

8: Try to work out farthest positions far left far right top or bottom etc. and remember. — DIMEN. IMAGE +ELIMINATION

9: No difference from last — DIMEN. IMAGE +ELIMINATION

12: Try as I read to slot people into a position from left to right <u>then drop middle names.</u> Lastly note who is above or below each other.	DIMEN. IMAGE +ELIMINATION
16: Remember people at extreme left + right and person above as a set of three names in that order.	DIMEN. IMAGE +ELIMINATION

3-term

17: A visual presentaton of Tallest at the top to smallest at the bottom	DIMEN. IMAGE
20: Solved the same way- remember only tallest and shortest- in that order.	DIMEN. IMAGE +ELIMINATION
24: Still a mental picture of height relationships pick out tallest and smallest	DIMEN. IMAGE +ELIMINATION
25: same	DIMEN.IMAGE +ELIMINATION
28: Same	DIMEN. IMAGE +ELIMINATION
32: Lost concentration- repeated relationship "shorter" made a <u>mental diagram more difficult to construct</u> as did additional name.	DIMEN. IMAGE

SUBJ 23

5-term

1: I worked out the problem by placing the names in a shape each sectioned off from the others and all spatially different.	GLOBAL IMAGE
4: If there were letters used in alphabetical order it might be easier to remember which position they were in, as sometimes it is easy to remember which one was at the bottom etc because their position on the screen but difficult to remember the names. I also find that when the question refers to the odd man out eg.	DIMEN. IMAGE

Bob to the left of Jim
Tom to the left of Bob etc
and then: John Below Bob

- This is easier than saying who was positioned to the right the most.

I like to arrange the names in my head in linear fashion.

8: I find that I read lines 1 and 2 and then search for a name that I have seen in these in line 4 (as line 3 is usually completely different) Line 4 links up 1 and 2 and lastly I read line 4.	SCANNING
9: I use mental imagery always - never vocal- I don´t seem to rehearse in my head because I can´t remember names well. The names can be associated with friends or enemies who have them. Eg. personal friend Mike- So I anticipated a question which would be answered with the name Mike ie. Who is rightmost and then when the question ´Who is leftmost´ came up I couldn´t answer it.	GLOBAL IMAGE OTHERS: MNEMONIC
12: There are two variations of the task. The easier of the two is the one where line 2 follows on from line 1 eg. Tom is to the left of Henry Jim is to the left of Tom	---
16: I feel I am improving at the task because I look for the line that follows on from line 1. This is either line 2 or line 4.	SCANNING

3-term

17: I visualised the three girls standing side by side in order of height.	DIMEN. IMAGE
20: Sometimes if the names can be put in alphabetical order according to their height I do this eg Bob - tallest Harry Jim	OTHERS: MNEMONIC

24: - DIMEN. IMAGE

25: - DIMEN. IMAGE

28: - DIMEN. IMAGE

32: Sometimes I use thinness of letters to remember tallness. OTHERS: MNEMONIC

eg. Jim is taller than Mary

letters are fatter therefore she looks shorter

EXPERIMENT: 2 Minimum Solution Paths and Transfer

Summary:

In this experiment, two groups of subjects were required to solve 6 ToH isomorphs, followed by the MC problem and the SR problem. One group- the shallow group- solved each of the 8 problems twice. The other group- the MSP group- solved each of the 8 problems repeatedly until they had produced a solution in the minimum number of moves possible. Half of each group were asked to provide written reports of how they solved the problems. They were asked to consider which problems or which parts of problems they found particularly difficult and why. They were asked to consider what they learnt about the problem that allowed them to overcome these difficulties, and whether they could formulate any useful hints that would enable others to solve the problems. They were not asked to recall specific moves or move-patterns, neither were they informed of the structural affinities that existed between the problems. They were not informed before hand, exactly when they would be asked to provide reports. In fact they were asked to provide reports after their 1st, 3rd, 6th and 7th problems.

The Categories:

The protocols were analysed with respect to 4 categories (see chapter 3: pp. 82-84 and tables 3.2.4 & 3.2.5). The first of these was the RECOGNITION OF STRUCTURAL RELATIONS. This takes two forms; whether the subject gave any report of recognising that the ISOMORPHS were related in some way, and whether they recognised that the NON-ISOMORPHS were different from the isomorphs in some way. As can be seen from the protocol reports this is an uncontentious category to evaluate, most subjects giving a fairly unambiguous report such as "very similar to the parking lots". It must be noted however that few subjects actually report explicitly what it is that makes the problems similar. They certainly do not describe this similarity in terms of state-space equivalence or any other structural abstraction. Note also that a report such as "same principle as the first", does not qualify for this category since it is not clear whether the subject means that the problems, or his solutions are constructed on the "same principle".

The second major category is STRATEGY REPORT. The sub-category of RANDOM/TRIAL AND ERROR was not difficult to score since protocols were only so-classified if the subjects actually mentioned trial and error or making moves at random. But note that on several occasions subjects report making moves at random at first, and later forming a plan. In these case subjects score in the ´trial and error´ category and the ´others´ category.

For the sub-category of STRUCTURALLY MEDIATED STRATEGY, there was a strict criterion. To be so classified, a report had to show evidence that the subject had recognised the recursive nature of the problems. Thus it was not adequate simply to say; "I built them up in a sequence in the middle slot" or "plan ahead two or three moves" or "need to get blue to road B first". While any problem solving strategy based on understanding the state-space invariances of the ToH must incorporate all of these aspects, these reports do not identify the characteristic feature of such strategies namely, that the problem is decomposible into sub-problems all of which are solved in the same way: In order to solve the 4-ring ToH for peg C you must first solve the 3-ring ToH for peg B and so on (cf Simon 1975). Compare the above cited reports with that of L.P; "got to get the last one to where you´re trying to get them all. "So you got to get the second to last one to the middle," and later "now I can see that the same principle applies throughout the stack..."

The third category- RECOGNISE TASK-SPECIFIC DIFFICULTIES- was something of a ´hotch-potch´ of reports. Essentially, anything which referred to a characteristic of a problem´s surface structure form qualified as a member of this category. Thus; "shape of boats [meant] I felt [they] needed to placed the right way up" and "more difficult...[because].. small cylinders were hidden", constitute good examples.

Finally, the fourth category- OTHERS- mainly consists of reports which indicate some form of strategic problem solving which was not structurally mediated and was not just trial and error. Thus there are many reports of "planning 2 or 3 moves ahead" (indication of some form of depth first search perhaps), and "I had to go through every move.." (exhaustive search) and so on.

Example Protocols:

REPORT	CATEGORY
Subj: M.J. (MSP Group)	
1: Computers	
No set moves, rather plan ahead two or three moves with a goal in mind. ie. number 4 to the bottom, next number 3 to the bottom and so on.	OTHERS: PLANNING +STRUCTURAL STRATEGY
3: Galleries	
Same principle as the first, only was more sure of planning ahead. I could not remember set moves but recognised them as correct when they came to mind.	OTHERS: PLANNING
6: Rockets	
Same principles as other. By now I can remember set moves. This game was more confusing as I had remembered moves as big to the last, medium to middle etc, and the reversal of size order added confusion and involved conversion. Also more difficult to plan moves because difficulty in "seeing ahead" ie. small cylinders were hidden. Also the problem was unrealistic.	RECOG. DIFFICULTY +OTHERS: PLANNING
7: Missionaries and Cannibals	
Seemed to be a quite different problem. Developed a plan, strategy with time. Difficulty was the problem seemed to	RECOGNISE DISSIM. OTHERS: PLANNING

be quite new and time needed to solve. Solved by planning ahead two moves. Seemed to be only one possible combination of moves at each stage.

SUBJ: C.S. (MSP Group)

1: Wedding cakes

Clear bottom of discs to move onto goal.	STRUCTURAL STRATEGY

3: Parking Lots

Lane C had to be left nearly clear, so only could be used temporarily in order to let A become reversed ie. yellow had to get to yellow [c?] whilst everything else was in B in right order. A was only one that that could use all 3 as it was blocking more than others.	--

6: Harbours

Similar to previous. Shape of boats: I felt needed to be placed right way up. Could almost sense when gone wrong. Difficulty sometimes in getting right ships to arrive at right place. Have got to get D and North unblocked by moving the others much in the same way as lorries start by moving B to either and progress from there.	RECOG. SIMILARITY +RECOG. DIFFICULTY

7: Missionaries and Cannibals

Problem of outnumbering whilst at bank- reluctance to move 2 people back seemed to egg on error. Once a pattern could be seen fairly simple- Ist stage easy as could only move cannibals, but difficulty after that.	RECOG. DIFFICULTY

SUBJ: L.P. (MSP Group)

1: Parking Lots

You could tell when you´d don´t it wrong but forgotten how I did it now. Need to get blue to road B first.	--

3: Computers

Unlike the rockets you can see these so it helps. Very similar to the parking lots. Got to get the last one out to where your trying to get them all. So you´ve got to get the second to last one to the middle.

RECOG. DIFFICULTY
+RECOG. SIMILARITY
+STRUCTURAL STRATEGY

6: Wedding Cakes

Same as before but now I can see that the same principle applies throughout the stack. Moving disks alternately (to alternate places) and then bringing them back together again.

STRUCTURAL STRATEGY

7: Missionaries and Cannibals

Don´t think its the same. Didn´t try to relate it because didn´t think it did. But perhaps I should have thought about it more. Important to remember you can bring the people back -you don´t have to leave them there.

RECOG. DISSIM.
+RECOG. DIFFICULTY

SUBJ: P.S. (NON-MSP Group)

1: Computers

I remembered how I started off the experiment from the first time although the first moves were trial and error. I also remembered that the first time I built them up in a sequence in the middle slot which is what I tried to do the second time.

OTHERS: MEMORY
+TRIAL AND ERROR

3: Harbours

After the first time I was able to see a pattern and form a ´plan´ which I followed the 2nd time. The first time, moves were made at random. I think I did develop a strategy.

OTHERS: PLANNING
+TRIAL AND ERROR

6: Rockets

Similar to previous ones. Initially moves made at random, organised plan evolved. Haven't remembered moves but do remember the start neccessary to solve the problem and try to work towards that. Might eventually develop a strategy. I could tell the position needed at the start.	RECOG. SIMILARITY +TRIAL AND ERROR +OTHERS: MEMORY PLANNING

7: Missionaries and Cannibals

Much harder than others. Pretty much trial and error until the last couple of moves where I could see some kind of strategy.	TRIAL AND ERROR

SUBJ: A.M. (NON-MSP Group)

1: Parking Lots

The first few moves I made at random, I thought I might have to begin again but actually I don't think it makes much difference in the first few moves. After that you could see there were moves that would get you stuck, but I wasn't particularly organised about what I did.	TRIAL AND ERROR

3: Computers

I thought it had things in common with both of the other two I did beforehand. You just had to remember what you were eventually aiming at and make sure not to block yourself so that you had a large number to move and nowhere to put it. You maybe had to think two moves ahead but I wasn't organised.	RECOG. SIMILARITY +OTHERS: PLANNING

6: Art Galleries

This seemed pretty similar to the other ones but since I got stuck this time and had to start again ....I'm not so sure! It was the same principle though, looking	RECOG. SIMILARITY OTHERS: PLANNING

ahead so that you didn´t make useless moves that would block what you wanted to do next. I discovered on the second go that I couldn´t make a different first move and still succeed. May be that was just me not the game though.

7: Missionaries and Cannibals

That was difficult!! I´d done puzzles like that before and I knew there was a strategy but I couldn´t remember what it was. Couldn´t even remember what I´d done when I tried it the second time. I had to go through every move before I did something to make sure that no missionaries would ´get eaten´ and still have someone to row the boat ...Each step was a one-by-one move- too much mental strain if you started looking ahead!

OTHERS: PLANNING

SUBJ: D.M. (NON-MSP Group)

1: Wedding Cakes

The first 3 or 4 moves were totally random until I grasped that it was necessary to think more than one move in advance to save myself getting into time wasting situations. I think 2 or 3 moves in advance would be sufficient. Once I had grasped this and done a few more constructive moves ie. they were intended to get me towards the goal and not just move the pieces around, it became easier because I knew what kind of moves I was looking for. So the second time round was easier, apart from one point in the middle where I lost the thread of my thought and had to start from the beginning again, ie. work out what my previous moves had been aimed at, and what I had planned my next moves to be. This obviously wasted time.

TRIAL AND ERROR
+OTHERS: PLANNING

3: Rockets

Solving of this puzzle by the lessons I learned in doing the first one (the second one was helped by that as well). For some reason, however, I was aware of taking longer and making more mistakes on this one. Perhaps it took me a while to realise that it actually was the same as the others, meaning that, to start with, I made a few random time wasting moves. Once I got into the routine of it, it was OK: I didn´t learn the moves by heart, I just learned the sequence necessary to move ahead and carried on using that. The second time around I am concious of making one mistake, after about 50 seconds. I realised immediately that I made a wrong move and moved the piece back to its original position. A hint would be: don´t panic! and perhaps use some time before starting to move to actually work out the way you are going to move them. Ie. plan ahead right from the start and not from about move 5 onwards, which I have been doing so far. It is tempting to get the "feel" of the pieces by moving them around a bit before you start actually making constructive moves, like a warm-up. This would apply to all problems and not just to this one however.

OTHERS: MEMORY
+RECOG. SIMILARITY
+TRIAL AND ERROR
+ OTHERS:PLANNING

6: Harbours

Again I learned from the earlier problems. At least I felt like I did, but the first time round, at least I made a lot of mistakes and took a lot of time. The reason for this was I didn´t listen carefully enough to the instructions and so started moving the pieces to places suitable for ending up in the east canal rather than the west. This probably wasted about 30 seconds and 10 moves. Perhaps I´m getting to the stage of having such a number

OTHERS: MEMORY
+RECOG. SIMILARITY

of similar problems to do that I'm blasé as far as listening to instructions is concerned: carelessness stemming from over-confidence and boredom.

7: Missionaries and Cannibals

The one before this was really easy but this one was very difficult for some reason. Probably because there were 6 pieces to deal with instead of 4. I was moving at random nearly all the time, except at the end of the first time and at the very end of the second time I worked out what I should have been doing. I couldn't remember this though: too complicated to learn in just one go, without taking longer. The illegal moves represent a failure to look ahead. I don't know why I didn't, but it slowed me up.

RECOG. DIFFICULTY
+RECOG. DISSIM.
+TRIAL AND ERROR