PROBLEM STRUCTURE AND
PROBLEM SOLVING BEHAVIOR

by

George F. Luger

Research Fellow

Department of Artificial Intelligence

University of Edinburgh

Scotland

Gerald A. Goldin

Assistant Professor

Graduate School of Education

University of Pennsylvania

Philadelphia Pennsylvania USA

ABSTRACT

One of the most interesting questions in the psychology of problem solving is the nature of the interaction between a problem's intrinsic structure and a problem solver's strategies or behaviors. The present paper suggests the use of techniques developed in research in mechanical problem solving to assist in formulating and illuminating this question. The authors also seek to develop a relationship between artificial intelligence methods and 'structuralist' theories of cognition by relating groups of symmetry transformations and 'conservation' operations.

Section I: Introduction

One of the most interesting questions in the psychology of problem solving is the nature of the interaction between a problem's intrinsic *structure* and the *strategies* or *behaviors* employed in attempting to solve the problem. Several studies have recently appeared addressing this subject: 1) the learning of mathematical structures such as the Klein Group or the Cyclic group of order four, Branca & Kilpatrick [1], 2) the study of analogy and transfer in related problem solving situations, Reed, Ernst, & Banerji [2], 3) the development of mechanical theorem provers both in equation solving, Bundy [3], and in elementary geometry, Gelernter [4] and Goldstein [5]. This research, although from diverse points of view, shares a common interest: understanding the effects of problem structure - for instance, a problem's possible subproblem and symmetry decompositions - on efficient problem solving.

This paper suggests techniques that may further formulating and illuminating this question. More ambitiously, the authors seek to develop a relationship between artificial intelligence methods and Piagetian or 'structuralist' theories of cognition.

Nilsson [6] has defined the *state space representation* of a problem as the set of distinguishable problem configurations or situations together with the permitted moves or steps from one problem situation to another. Thus the state space of a problem consists of an initial state, together with all the states that may be reached from the initial state by successive legal moves in the problem. One or more of these successor states are classified as goal states. The state space of a problem, represented as a non-directed graph, will be *unique* only if the problem's description clearly deliniates its initial and goal state(s) and its set of legal moves. Finally, the concept of the state space of a problem can be generalized to the analogus structure for an N-player game, i.e., the game tree or graph.

Banerji [7], Banerji & Ernst [8] and others researchers have offered mathematical descriptions to characterize state spaces. This 'state space algebra' allows such concepts as problem comparison, decomposition, and extension to be well defined and also allows problem solving studies in the areas of problem analogy, transfer, and generalization to be extremely precise.

In early artificial intelligence research both Gelernter's Geometry Theorem Prover [4] with its use of the symmetries within the syntax of a problem's description, as well as Newell, Shaw, and Simon's General Problem Solver [9], with its utilization of a problem's possible subproblem decompositions affirm the need for as complete as possible exploitation of a problem's structure for effective problem solving. Again, in robot plan formation Sacerdoti [10] uses ABSTRIPS to focus on the important features of a problem's structure and to ignore the unnecessary detail that leads STRIPS to combinatorial problems.

Newell and Simon's later work [11] permits in principle a very detailed interpretation of an individual's problem solving 'protocol' as steps in information processing. However, as the 'problem space' for this research varies from subject to subject for each individual problem it also lends to their model a definite post hoc character. Since no final commitment concerning the structure of the 'problem space' is made until after the problem solving is observed, the potential for predicting the effects of a problem's structure on a subject's problem solving behavior seems to be lacking.

In the next section two ideas are introduced. First, we assert a fundamental correspondence between conservation operations and symmetry transformations. In the sense of Piaget, a conservation operation is the ability of a problem solver to respond that two different states of the environment are equivalent when they are functionally the same, that is when they both possess the same value for some perceptual or cognitive variable. For example, 17 is said to be equivalent to 32 modulo 3, since both have the same remainder on division by 3. In general, a symmetry transformation is a mapping which carries one problem state into another in such a way as to leave unchanged important observable features. In the everyday sense of the word symmetry these features are geometric, for example, the transformation which changes a particular configuration of objects into its 'mirror image' may leave the appearance of the configuration unchanged. We are interested, however, in a more general symmetry, for example, that in problem descriptions and underlying problem structure, as well as in the readily apparent geometric symmetries.

The second idea pursued in this section is that in problem solving, a subgoal and subproblem decomposition of a problem may govern a problem solver's behavior even when he or she is not consciously seeking to arrive at that particular subgoal, and despite the fact that the infrastructure of subproblem's within the main problem may not on the surface be apparent. Furthermore, given a subproblem decomposition, one kind of symmetry whose effect may be explored is the presence in the problem of subproblems of identical (isomorphic) structure.

In the third section additional concepts concerning a problem's state space are rigorously defined, and several hypotheses offered concerning effects of problem structure on subject's paths through the state space, for example a predominance of goal and subgoal directed paths, and an increased likelihood of congruent paths through isomorphic subproblems.

In the fourth section, the Tower of Hanoi problem Nilsson [6] and the Tea Ceremony problem Newell [12] are used to illustrate the main ideas developed. Finally, some suggestions for further experimental investigation are proposed.

Section II:

(A) Conservation Operations and Symmetry Transformations

(B) Subproblem Decompositions

(A) In Tic-Tac-Toe (Noughts & Crosses) a player, say X, is said to 'fork' his opponent when he places his X in such a position on the board that 1) there are as a result two possible 'winning moves' for X, and 2) O is able, in the next move, to block only one of these 'winning moves'. There are several different 'forking positions' possible on the Tic-Tac-Toe board (see Figure 1). These 'forking' relationships are conserved or invariant over all rotations and reflections of the game board, and so it can be said that there is a conservation or functional equivalence among the different 'forking' situations. It is also possible to construct symmetry transformations or mappings of one forking situation onto any other. The authors would like to establish a logical equivalence between conservation operations and groups of symmetry transformations in characterizing this and other problem solving situations.

The correspondence between conservation laws and symmetries of nature is well known in modern physics. Conservation of momentum derives from the invariance of physical interactions under spatial translations, conservation of angular momentum from rotational invariance, conservation of energy from invariance under time translations, etc. Feynman [13].

That such a correspondence existed as a general principle long remained unobserved. In some instances physicists became aware of and successfully expressed a conservation law prior to understanding that the law actually derived from a known symmetry of the physical world, for example, in the cases of conservation of momentum, angular momentum, and energy. In other instances the symmetry was well known, and physicists proceeded to define an observable whose conservation followed automatically from the fact of obedience to the symmetry. Thus conservation of parity follows from the supposed invariance of physical interactions under spatial reflection. Such newly defined observables proved immeasurably useful when it was learned that on a sub-atomic level, symmetries such as spatial reflection which had heretofore been taken for granted were subject to violation, and non-conservation occurred. Likewise there were some well known conservation laws, based on which previously unknown symmetries could be defined. Thus conservation of electric charge can be interpreted as a consequence of invariance under rotations in an abstract mathematical space (Isotopic spin-space [14]).

Today it is understood that the pairing of a conservation law with a symmetry may be regarded as a mathematical rather than an empirical relationship, which follows from the mathematical theory of Lie groups (Feynman [13]). This relationship asserts that to every set of observables corresponds a certain algebra of observables; and to every such algebra corresponds a group. If the values of the observables are conserved, i.e., unchanged as the system develops in time, then it turns out that the group elements describe physical symmetry transformations of the system. Further, any pair of symmetry transformations may be performed successively to generate a third symmetry transformation, defining an associative binary operation. The identity transformation is included as a symmetry by convention, and to every symmetry transformation there corresponds the inverse transformation which returns the system to its original configuration (Feynman [13]).

The group is the paradigm in mathematics of the methodology which has been termed 'structuralist' (Piaget [15]). This methodology has been applied to fields of study as diverse as anthropology, linguistics, and psychology, as well as to mathematics [15]. According to Piaget a structure in the most general sense is a system or set within which certain relations or operations have been defined, embodying the concepts of wholeness, transformation, and self-regulation. For example a system of kinship constitutes a structure in anthropology as does a group in mathematics. In Piagetian developmental psychology, the conservation operations - conservation of number, volume, quantity, etc. - are transformations which represent the cognitive structures assumed to underlie certain patterns of behavior. Acquisition of these conservation operations by children defines sequential stages in their cognitive development.

In view of the parallel fundamental roles played by group structures in mathematics and cognitive structures in developmental psychology, it is natural to try to look at the acquisition of Piagetian conservation operations as equivalent to the acquisition of a group of symmetry transformations.

For an observable (such as number, quantity, etc.) to be conserved means in fact that when a given state is somehow transformed into an altered state, the value of the observable is unchanged from its initial value. Of course, for the second state to be regarded as different from the first state at all, there must be at least one other observable which does change in value under the transformation, and which is not conserved by the transformation. A symmetry transformation may be defined, then, as a one-to-one mapping from the set of states onto itself which leaves invariant the specified relationships among the states. Any collection of such symmetry transformations generates a symmetry group.

Let us say that a certain symmetry group G conserves a given set of observables when for each state S in the system, all states which may be obtained from S by applying symmetry transformations from G have exactly the same values of the specified observables. The maximal symmetry group possessing this property for a given set of observables is the group containing every symmetry transformation which preserves the values of the specified observables.

As an example, consider the rearrangement of n objects on a table or two-dimensional surface depicted in Figure 2. The final configuration of objects (described by the coordinates x_1 .... x_n') may be obtained from the initial configuration (x_1 ... x_n) by means of a rearrangement mapping which appropriately transforms the points in the two-dimensional plane. Such a rearrangement must be one-to-one (so that two objects do not end up at the same point) and reversible. Since any two rearrangement mappings may be applied successively to yield a third, the set of all such mappings forms a group K. For this example the collection of states is the set of all possible configurations of n objects on the two-dimensional surface, for n = 0, 1, 2 ... .

To say that a subject 'conserves number' means that no matter how a given state of the environment is transformed into an altered state simply by moving the objects around, the value of the observable 'number' - according to the subject's report - remains unchanged. The group K defined above, that is the group of one-to-one surjective mappings from a region of R^2 onto itself, maps the set of states onto itself in such a way that a state specified by n points continues to be specified by n points after it is transformed, and thus has the same value of the observable 'number'. It is not difficult to see that K fits the definition of a symmetry group conserving that observable. Thus the acquisition of 'number conservation', that is the ability of a subject to respond that the number of objects remains unchanged when only the positions of the objects have been changed, is logically equivalent to the acquisition of the structure of the symmetry group K, that is, the ability to undo (invert) any rearrangement transformation and to catenate any two such transformations successively.

It may be hypothesized that stages in the acquisition of such a symmetry group structure actually correspond to the acquisition of particular subgroups of this symmetry group. For example, a child might at some time respond consistently that the number of objects is unchanged when a configuration is merely translated a certain distance in space, without its having been spread out or otherwise rearranged. If this were to occur we would say that the subgroup of K containing all translations had been acquired as a symmetry structure. Verification of this hypothesis would further demonstrate the usefulness of the conservation operation/symmetry group correspondence.

In arguing for the reformulation of conservation operations in terms of symmetry groups, it seems natural to cite examples of systems in which the symmetries are familiar, but the identification of conserved quantities may be cumbersome. Many examples drawn from problem solving turn out to be easier to describe in terms of symmetry groups than in terms of quantities conserved by the transformations in those groups. For example, Tic-Tac-Toe or Noughts and Crosses. In this game there are nine distinguishable states which can be reached by the first move of the first player. However, modulo the rotation or reflection symmetry, only three distinguishable states exist (see Figure 3). In constructing the state space representation for Tic-Tac-Toe, one could choose to represent all the distinguishable states of the system, and so obtain a very large state space; or one could use the much smaller state space obtained by regarding those states conjugate by symmetry as equivalent. This latter choice corresponds to reduction of the state space representation modulo its symmetry transformations

In studying human problem solving, we must take into account the possibility that the subject's behaviour does not initially reflect all the symmetry which is actually present. Therefore, to map the subject's behaviour faithfully, we should begin with the expanded state space representation i.e. the state space containing all possible legal states of the problem. This expanded space (and its formal properties) will be constant across all subjects solving this problem and so eliminate post hoc analysis (p.2)

Tic-Tac-Toe provides an example of a game in which the rotation and reflection symmetry is easily recognized, but the corresponding conserved quantities are cumbersome to define. For example, one such quantity would be the number of X's in corner squares, a number unchanged by the rotation or reflection operations. Number Scrabble [11], a game isomorphic to Tic-Tac-Toe, may be described as follows. The integers 1, 2, 3, ... , 9 are written on a pad, and the two opposing players take turns selecting single numbers for himself. Neither player may select a number already taken. The goal is to obtain any three numbers which add up to exactly fifteen. Figure 4 illustrates the isomorphism between this game and Tic-Tac-Toe. A player trying to learn this game would not have available the geometric symmetry presented by the Tic-Tac-Toe grid. Without prior faciliarity with the magic square, a player would have to seek rules such as, 'If the first player chooses 5, then the second player must pick an even number to avoid losing'. Unbeknownst to the player, the relevant 'observables' are just those which are conserved by the Tic-Tac-Toe symmetry __ 'even numbers selected', 'odd numbers excluding 5', and so on. Tic-Tac-Toe and Number Scrabble illustrate (a) that symmetries may be more convenient than the quantities conserved by those symmetries for formulating the notion of equivalence among states, (b) that symmetries and conserved quantities are, however, logically interchangeable, and (c) that the rules of a game may be reformulated in such a fashion as to make identification of the conserved quantities easier or more convenient than the characterization of the symmetries. Finally, the formal correspondence between a group of symmetry transformations and the observable quantities conserved by these symmetry transformations suggests that acquisition of symmetries may be as fundamental to cognitive development as is the acquisition of conservation operations. We have also seen how the presence of symmetry may be represented in the state space of a problem or game.

(B) A second feature of a problem which is amenable to study utilizing the state space is a problem's infrastructure of subproblems. It has been commonly held that an effective problem solving technique is to establish subproblems or subgoals whose solution or attainment might assist in the conquest of the main problem. Polya [16] suggests such an approach in discussing his problem solving 'heuristics', it forms the basis of Newell, Shaw, & Simon's General Problem Solver [9], and suggests to Nilsson [6, p 80] one way to reduce the state space. But to establish rigorously the role of such identification of subgoals in human problem solving behaviour remains difficult and psychologists are divided even over the assumption of 'goal-directedness' (Kimble [17, sec. 13]). Characterization of subproblems as subspaces of the problem's state space should assist in investigating the behavioral consequences of a subproblem decomposition by the problem solver. One may further discuss, independently, the group of symmetry transformations of a subproblem, or explore the effects of the presence in a problem of different subproblems having identical (isomorphic) structure.

The above considerations suggest the utility of mapping the problem solver's steps as paths through the state space representation of the problem. Based on the formal properties of the specific problem's state space, such as its symmetry and decomposition into subproblems, hypotheses can be formulated which predict the effect of this structure on the paths generated by the problem solver. Then the door is open to the development and empirical test of general algorithmic or mechanical procedures that might replicate the properties of the paths generated by human problem solvers. The decision to represent problem solving behavior as paths through the state space of the problem is motivated by the desire to make precise the data which needs to be 'explained' by a theory of human problem solving.

In practice it may not always be easy to represent behavior in this fashion, since the uniquness of a problem's state space representation relies on the preciseness of the problem's statement. Further, a problem solver's production of paths depends on his or her ability to discriminate among the perceptual or cognitive variables which characterize the states and legal moves of the problem. The best experimental situation then, is a problem whose states correspond to different discrete situations of an actual physical device, such as Tic-Tac-Toe, N-pile NIM, or the Tower of Hanoi and Tea Ceremony problems to be discussed in section four. Other available means for recording a subject's behavior as a succession of states entered may include recordings of oral comments, written notes, or even gestures and eye movments (Bartlett [18], Newell & Simon [11], and Young [19]).

Section III: (A) Definitions and (B) General Hypotheses

(A). Before proceeding with further discussion, definitions are given for the concepts central to the present approach. These definitions are based on and expanded from those given by Nilsson [6]. The state space of a problem is the set of distinguishable situations or states of the problem, together with the permitted transitions or moves from one state to another. The problem must specify an initial state and one or more goal states, and so the state space may be visualized as a non-directed graph (Figure 6).

A subspace of the state space is a subset of the states, together with the permitted transitions which obtain between these states in the subset. A subproblem is a subspace of the state space with its own initial and subgoal(s) states. For a subproblem it is required that if the initial state is not the initial state of the problem, it can be entered from a state outside the subspace; and if a subgoal state is not a goal of the main problem, it can be used to exit from the subspace - i.e., to enter a state of the problem outside of the subproblem. There are often many ways to decompose a particular problem into subproblems, which correspond to different choices of subspaces within the state space.

Two problems (or subproblems) are said to be isomorphic if and only if there is a bijective mapping from the state space of the first onto the state space of the second and: 1) the initial state of the first problem is mapped onto the initial state of the second, 2) the set of goal states of the first problem is mapped surjectively onto the goal states of the second, and 3) a transition from one state to another is permitted in one problem if and only if the corresponding transition is permitted in the other.

An automorphism of a problem is an isomorphism of the problem onto itself and is called a symmetry transformation or symmetry automorphism. The set of all the automorphisms of a problem forms a group under the binary operation of composition or the successive application of two automorphisms. This group is called the symmetry group or automorphism group of the problem.

The states of a problem may be distinguished by virtue of having different discrete values for a set of variables called observables. These observables, characterizing the problem states, may refer to color, position, or number, etc. An observavle is said to be conserved by a group of symmetry transformations, if and only if for any state, the value of that observable is unchanged by any element of the group of transformations.

Let S be a state of a problem, and consider the set of all states which can be obtained by applying automorphisms or symmetry transformations from a group G to S. This set of states is called the orbit of S under the automorphism group G. Two states are said to be conjugate modulo the symmetry group G if they are in the same orbit under G.

The orbits within the state space form mutually disjoint equivalence classes of states. A new and simpler state space may now be constructed canonically by considering each equivalence class as a state in its own right, or alternatively, by selecting one representative state from each orbit. The state space thus obtained is said to have been reduced with respect to its symmetry group G, or reduced modulo G. G may be the full automorphism group of the original state space, or any subgroup thereof..

A path in the state space of a problem is a sequence of states $S_1, S_2, \ldots, S_n$ such that for $i = 1, 2, \ldots, n-1$, the pair S_i, S_{i+1} represents a permitted transition of the problem. A solution path for a problem is a path in which S_1 is the initial state and S_n is a goal state, with $S_2, \ldots, S_{n-1}$ neither initial nor goal states of the problem. Two paths within respective isomorphic problems are said to be congruent (modulo the isomprphism) if one path is the image of the other under the isomorphism.

We have seen above that one way to reduce the size of the state space is with respect to a group of symmetry automorphisms of the problem. A second means of state space reduction is with respect to the subproblem structure. The state space may be described, albeit nonuniquely, as a union of mutually disjoint subspaces, such that for any ordered pair of subspaces, a transition exists from a state in the first to a state in the second. An entire subspace may thus be regarded as a single state in the reduced state space, and a transition is permitted from one subspace to another whenever a transition does in fact exist from a state in the one to a state in the other. Each subspace, now a state in the reduced state space, becomes also a subproblem of the original problem whenever a particular entry state is designated as 'initial', and any or all of its exit states are designated as 'goals'. We then say that the state space has been reduced modulo its subproblem decomposition.

Finally, one may address the concept of a non-random or a goal-directed path within a problem or subproblem. Roughly speaking, a non-random path would differ locally - perhaps in the number of 'turns' or 'loops' - from random paths generated through a problem's state space representation. A goal-directed path is a solution path which does not 'double back' on itself within the state space, moving consistently 'towards' rather than 'away from' the goal state. Criteria for defining 'loops', or 'doubling back', or 'distance from the goal state', etc, are for the present to be established in the context of each specific problem under consideration. While these criteria may differ across problems of different structure, they will remain constant across populations of subjects solving a particular problem.

(B) In problem solving it may be assumed that the solver acts sequentially upon problem situations (states) to generate successor states, a process which can be described, as discussed above, by means of paths through a state space representation of the problem. It is nowhere suggested that the problem solver 'perceives' the state space as an entity during problem solving. The symmetry properties which have been discussed are formal properties of the state space, which may (as in Tic-Tac-Toe) or may not (as in Number Scrabble) correspond to geometrical or perceptual properties of the problem readily apparent to the problem solver.

The approach to this stage of research has been to formulate hypotheses respecting the paths generated by problem solvers in the state space of a problem. Such hypotheses 1) are motivated by the formal properties of the state space under discussion, and 2) represent the anticipated effects of the problem structure in shaping problem solving behavior. The following hypotheses of a more-or-less general nature are suggested.

Hypothesis 1. (a) In solving a problem (or subproblem) the subject generates non-random, goal-directed paths in the state space representation of the problem (or subproblem), and (b) when sub-goal states are attained, the path exits from the respective subproblems.

Hypothesis 2. Identifiable 'episodes' occur during problem solving corresponding to the solution of various subproblems. That is, path segments occur during certain episodes which do not constitute the (direct) solution of a problem, but which do constitute the solution of the isomorphic subproblems of the problem.

Hypothesis 3. The problem solver's paths through isomorphic subproblems tend to be congruent.

Hypothesis 4. Given a symmetry group G of automorphisms of the state space of a problem, there tend to occur successive path segments congruent modulo G in the state space.

It may be that the validity of hypotheses 1 and 2 depends on the particular way that the state space of teh problem is decomposed into subproblems since such a decomposition is often not unique. Hypothesis 4 (symmetry acquisition) is suggestive of the 'insight' phenomenon which changes the gestalt of the problem solver (Wertheimer [20]) and often plays an important role in the eventual problem solution.

These hypotheses are not to be regarded as a definitive list, but rather as preliminary and indicative of the kind of analysis possible of the effects of problem structure on the problem solver's behavior. If valid, these hypotheses would offer fairly general constraints on the properties which mechanical models must display to simulate human problem solving.

Section IV:
Two Problem Solving Studies and Suggestions for further Research

Let us seek to make the foregoing ideas more concrete by considering two problems that have been used for empirical investigation (Luger [21]). The Tower of Hanoi problem has been extensively discussed in the literature and its state space considered by Nilsson [6]. It is a natural problem to consider both because its well defined state space has a rich subproblem structure and because its state space possesses somewhat more symmetry than is immediately apparent in the problem environment.

In the Tower of Hanoi problem four concentric rings (labelled 1, 2, 3, 4 respectively) are placed in order of size, the largest on the bottom, on the first of three pegs (labelled A, B, C); the apparatus is pictured in Figure 5a. The object of the problem is to transfer all the rings from peg A to peg C in the minimum number of moves. Only one ring may be moved at a time, and no larger ring may be placed over a smaller one on any peg.

The Tea Ceremony, see Figure 5b, is an isomorph of the Tower of Hanoi. Three people - a host and an elder and younger guest - participate in the ceremony. There are four tasks they perform - listed in ascending order of importance: feeding the fire, serving cakes, serving tea, and reading poetry. The host performs all the tasks at the start of the ceremony, and the tasks are transferred back and forth among the participants until the eldest guest performs all the tasks, at which time the ceremony is completed. There are two constraints on the one-at-a-time transfer of tasks: 1) only the least important task a person is performing may be taken from him, and 2) no person may accept a task unless it is less important than any task he is performing at the tims. The object of the Tea Ceremony game is to transfer all the four tasks from the host to the elder guest in the fewest number of moves. As with the Tower of Hanoi, the subject attempts the game repeatedly, starting over again whenever he or she wishes until the rings are moved (or tasks transfered) in the fewest possible number of transitions.

In the isomorphic relationship between the Tea Ceremony and the Tower of Hanoi the people - host, youth, and elder - correspond respectively with pegs A, B, and C. The four tasks - feeding the fire, serving cakes, serving tea, and reading poetry - correspond respectively with rings 1, 2, 3, and 4. It can be checked that the initial state, goal state, and legal moves of the two games correspond.

Figure 6 is the complete state space representation of the Tower of Hanoi/ Tea Ceremony problem. Each circle stands for a possible position or state of the games. The four letters labelling a state refer to the respective pegs (people) on which the four rings (tasks) are located. For example, state CCBC means that ring 1 (fire), ring 2 (cakes), and ring 4 (poetry) are in their proper order on peg C (performed by the Elder). Ring 3 (tea) is on peg B (performed by the youth). A legal move by the problem solver always effects a transition between states represented by neighboring circles in Figure 6. The solution path containing the minimum number of moves consists of the fifteen steps from AAAA to CCCC down the right side of the state space diagram.

The Tower of Hanoi/Tea Ceremony has a natural decomposition into nested subproblems. For example, to solve the 4-ring Tower of Hanoi problem, it is necessary at some point to move the largest ring from its original position on peg A to peg C, but before this can be done the three smaller rings must be assembled in their proper order on peg B. The problem of moving the three rings from one peg to another may be termed a 3-ring subproblem, and constitutes a subset of the state space of the 4-ring problem. The 4-ring state space contains three isomorphic 3-ring subspaces, for which the physical problem solving situations are different by reason of the position of ring 4.

Each subspace becomes a subproblem when one of its entry states is designated as the initial state, and its exit states are designated as goal states. Similiarly, each 3-ring subspace contains three isomorphic 2-ring subspaces for a total of nine in the 4-ring state space; and each 2-ring subspace may be further decomposed into three 1-ring subspaces, comprising only three states apiece. Note the examples in Figure 6 of 1-, 2-, and 3-ring subspaces.

Each n-ring subproblem, as well as the main problem, admits of a symmetry automorphism. The automorphism maps a goal state of the n-ring problem onto the conjugate goal state which corresponds to transferring the m rings to the other open peg. Were the three pegs of the Tower of Hanoi board to be arranged at the corners of an equilateral triangle (as are the people in the Tea Ceremony), the symmetry automorphism would represent the the geometric operation of reflection about the altitudes of an equilateral triangle.

Criteria are established [21] for 'non-randomness' and 'Goal-directedness' of subject's paths through the Tower of Hanoi/Tea Ceremony state space. The number of 'turns' and 'loops' of a subject's path is compared with the 'turns' and 'loops' of the same length generated in the Tower of Hanoi/Tea Ceremony state space. A 'metric' is defined also, the function of the number of states the subject's current state is distant from the goal state. If this function is non-increasing over the subject's path, the path is said to be 'goal-directed.' This same metric is established to measure goal directedness within subproblems. When subgoal states are attained the path that exits from the subgoal is examined to see if it also exits from the subproblem. Figure 7a shows a subject's path that decomposes the state space modulo its 2-ring subproblems - each 2-ring subproblem is solved in the minimum number of steps, while the 3-ring subproblem is not. This represents a 2-ring 'episode' in the problem's solution. Figure 7b shows three congruent paths through isomorphic 2-ring subproblems.

The problem solving data of 45 adult subjects solving the Tower of Hanoi and 21 adult subjects solving the Tea Ceremony problems are reported by Luger [21]. Except for Hypothesis 3 (the production of congruent paths through isomorphic subproblems), all the hypotheses are supported by the data. Especially strong (near 100 %) is the support of the special role played by subgoal states within the problem (Hypothesis 1b). 86 % of all the subjects have at least one problem solving 'episode' with 60 % showing two or more of the three theoretically possible 'episodes' (Hypothesis 2). 52 % of all subjects in the studies interrupted a path and immediately produced a path segment that was the symmetric conjugate of the interrupted path (hypothesis 4). This new path was often the minimum step solution path.

Figure 8a pictures the actual paths through the state space generated by one adult subject solving the Tower of Hanoi problem. This subject's behavior happened to conform to all four proposed hypotheses. The paths are both goal- and subgoal-directed, and exit from the subproblem whenever a subgoal state is entered. The first two trials contain five instances out of seven of minimum solution of the 2-ring subproblem, while the 3-ring subproblem has not yet been solved by the shortest path - a 2-ring 'episode'. Trial 1 illustrates two congruent non-minimum paths through 3-ring subproblems. Finally, trial 2 is interrupted and trial 3, the shortest solution path, follows as the image of trial 2 under the symmetry automorphism that exchanges pegs B and C.

Figure 8b pictures the paths of an adult subject solving the Tea Ceremony problem. The paths within each problem are goal-directed, and whenever a subproblem's goal state is entered it is left by the unique path that also leaves the subproblem space. From the beginning the problem is reduced modulo its 2-task subproblems, since during the problem solving 12 of 14 of the 2-task subproblems are solved in the minimum number of steps. After the first 3-task subproblem, there is an 'episode' in which 5 of 6 of all further 3-task subproblems are solved in the minimum number of steps but the entire problem (4-task) is not, reducing the problem by its 3-task subproblems. The first and second trials begin with congruent paths (non-minimum) through 3-task subproblems. The third trial is interrupted, and its symmetric conjugate - which solves the problem- is produced in the fourth trial.

In summary, the present paper suggests one natural way to make the strategy/structure distinction. We let the structure of a problem refer to the formal properties of its state space representation, such as its symmetry automorphisms and possible subproblem decompositions. We consider the subject's possible cognitive structures to include the conservation operations, symmetry and subproblem decompositions that the subject can apply to the problem situation. An example of this in the Tower of Hanoi is the ability of a subject to solve all 2-ring subproblems, no matter where they are in the in the context of the problem, in the minimum number of steps. These structures determine the states that the subject treats as distinct and those treated as equivalent. These may change during problem solving, leading to an effective reduction of the state space. A subject's behavior may be faithfully mapped as long as the state space representation that is utilized by the researcher is sufficiently detailed, in that it does not treat states as equivalent which the subject treats as distinct.

We let the term strategy refer to particular rules or procedures for taking steps within the state space. Different individuals may employ different strategies in solving the same problem, and the same individual may employ different strategies in solving different but isomorphic problems. The present paper does not explain strategies *per se*, but hypothesizes that even in the context of different strategies, certain patterns of behavior tend to occur as a consequence of the structure of the problem.

There are several obvious, very broad directions for further experimental research, including: 1) the test for 'transfer' in the behavior of a subject solving different problems having related - isomorphic or homomorphic - structure. The first author has in fact a study in progress looking for transfer across isomorphic problem situations, and Reed, Ernst, & Banerji [7] have examined homomorphic problems; 2) the occurance of developmental stages or substages corresponding to the acquisition of symmetry groups or their subgroups; 3) age, sex, or cross-cultural differences in the effects of problem structure on problem solving behavior.

References

1. Branca, N. & Kilpatrick, J. 'Consistency of Strategies in Learning Mathematical Structures', Journal for Research in Mathematics Education, Vol 3, No 3.
2. Reed, S.K., Ernst, G.W., & Banerji, R.B., 'The Role of Anology in Transfer Between Similiar Problem States', Cognitive Psychology, 6, 436-50 ('74).
3. Bundy, A., 'Analysing Mathematical Proofs', DAI No. 2, Department of Artificial Intelligence, University of Edinburgh, SCOTLAND.
4. Gelernter, H., 'A Note on Syntactic Symmetry and the Manipulation of Formal Systems by Machine', Information and Control, 2, 80-89, ('59).
5. Goldstein, I., 'Elementary Geometry Theorem-Proving', Artificial Intelligence Memo No. 280, M. I. T., USA.
6. Nilsson, N. Problem Solving Methods in Artificial Intelligence, New York: McGraw-Hill 1971.
7. Banerji, R.B., Theory of Problem Solving, New York: Elsevier Pub. Co. 1969.
8. Banerji R.B., & Ernst, G.W., 'Strategy Construction Using Homomorphisms Between Games', Artificial Intelligence, 3, 223-49 (1972).
9. Newell, A., Shaw, J.C., & Simon, H.A., 'Report on a General Problem Solving Program', RAND Corporation paper P-1582 (1959).
10. Sacerdoti, E.D., 'Planning in a Hierarchy of Abstraction Spaces', 3IJCAI, Advance Papers on Conference, California: Stanford Research Institute.
11. Newell, A. & Simon, H.A., Human Problem Solving, New Jersey: Prentice-Hall 1972.
12. Newell, A. Private Communication.
13. Feynman, R.P., Lectures on Physics, Vol III, Mass: Addison-Wesley 1965.
14. Eisenbud, L. & Wigner, E.P,, Nuclear Structure, New Jersey: Princeton Press 1958.
15. Piaget, J., Structuralism, New York: Basic Books 1970.
16. Polya, G., How To Solve It, New Jersey: Princeton Press 1945.
17. Kimble, G.A., Conditioning and Learning, New York: Appleton-Century-Crofts 1961.
18. Bartlett, F.C. Thinking, New York: Basic Books 1958.
19. Young, R.A., Production Systems as Models of Cognitive Development, Bionics Research Report No. 22, Dept. of A. I. University of Edinburgh, SCOTLAND.
20. Wertheimer, M., Productive Thinking, New York: Harper & Bros 1945.
21. Luger, G.F., PhD Thesis, University of Pennsylvania, Philadelphia, USA, 1973.

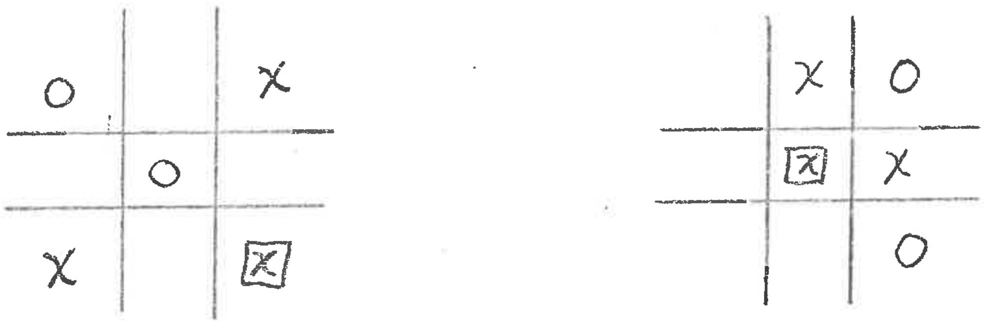

Figure 1. Two different 'forking moves' on the Tic-Tac-Toe board. The 'forking move is indicated [X] and it is O's turn to play. Each of the forks is conserved or invariant over the rotations and reflections of the board, and symmetry transformations exist mapping one fork onto any other.

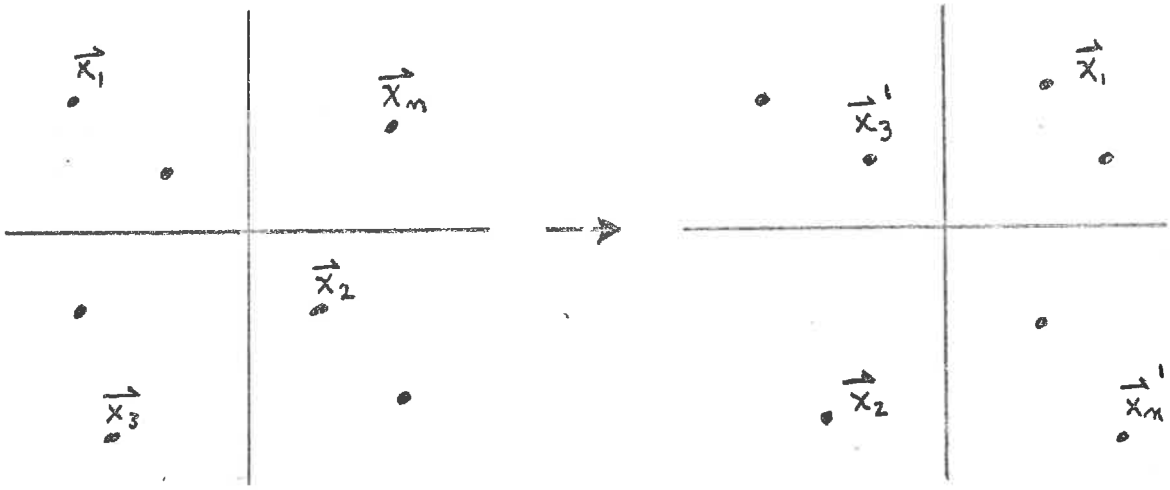

Figure 2. A rearrangement of n objects in two dimensional space. This transformation may be implemented by means of a one-to-one reversible rearrangement mapping.

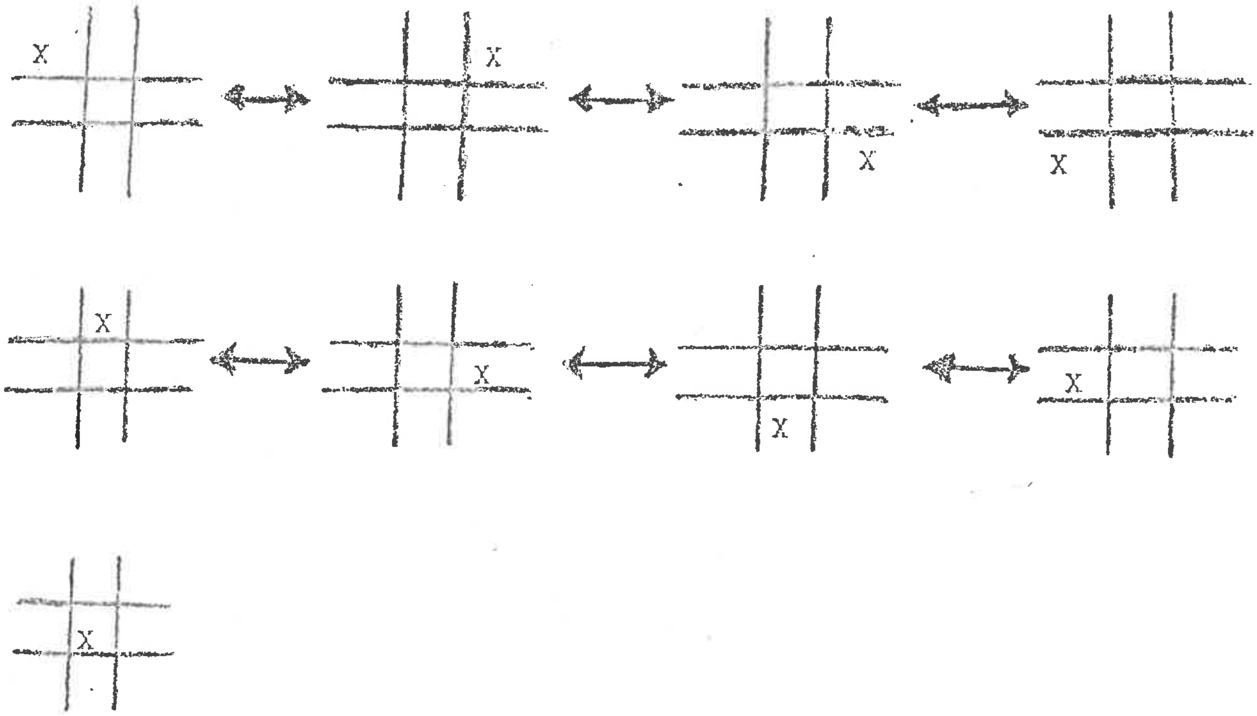

FIGURE 3. Tic-Tac-Toe states equivalent by symmetry.

4	3	8
9	5	1
2	7	6

FIGURE 4. Number Scrabble using integers 1, 2, ... , 9.

This illustrates the isomorphism between the number selection game described in the text and Tic-Tac-Toe.

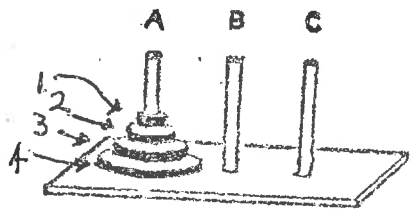

Figure 5a. The Tower of Hanoi board in its initial state.

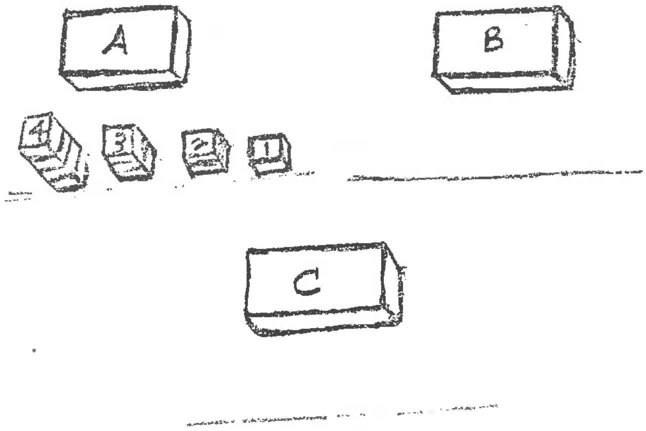

Figure 5b. The Tea Ceremony in its initial state. A,B,C, 1,2,3,4 show the isomorphism. A,B,C represent the Host, Youth, and Elder. 1,2,3,4 represent feeding the fire, serving cakes, serving tea, and reading poetry. The size of the block represents the order of importance of the task.

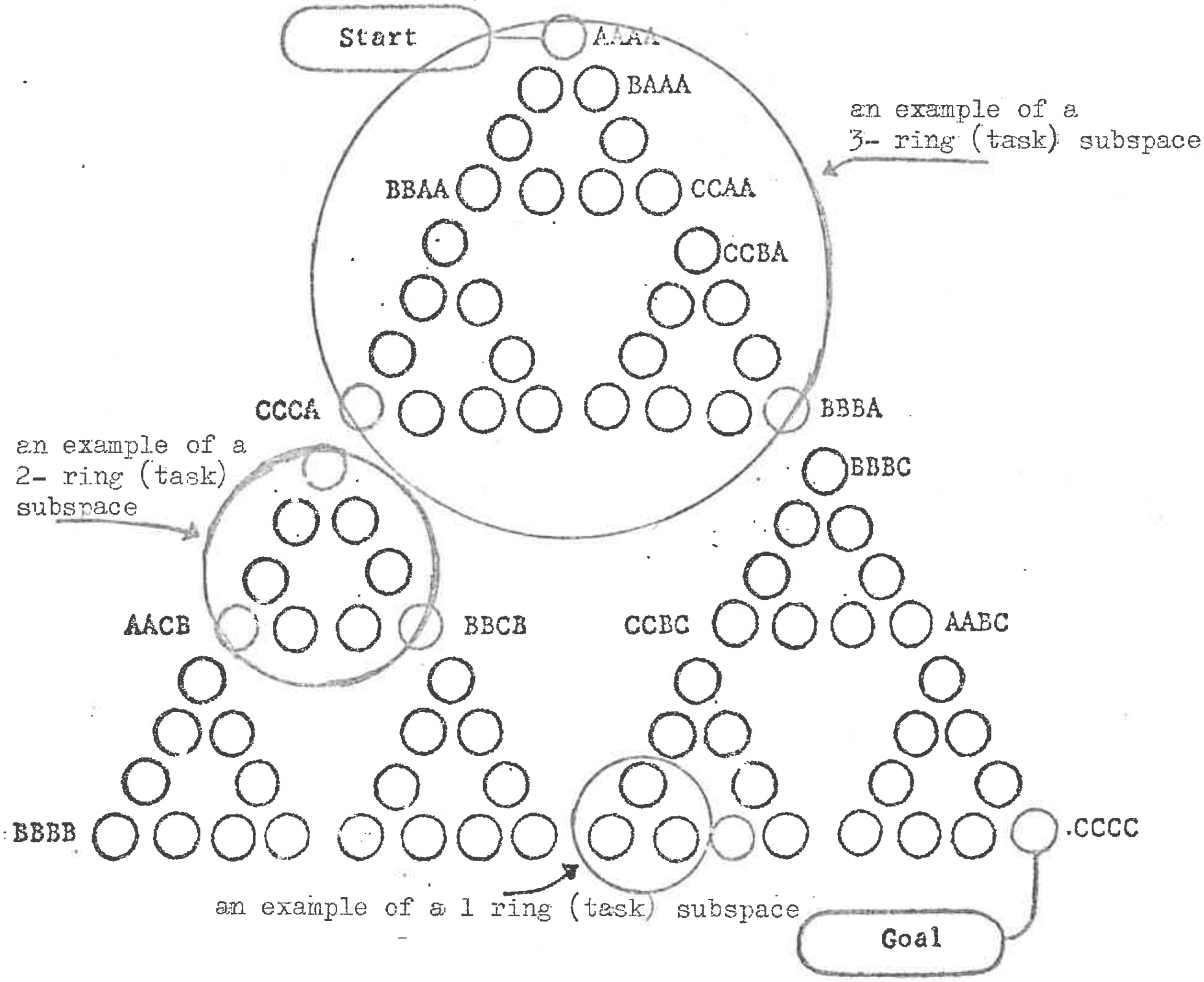

Figure 6. State space representation of the Tower of Hanoi/ Tea Ceremony Problem

The four lettere labelling a state refer to the respective pegs (persons) on which the four rings are located (by which the four tasks) are preformed. Legal moves effect transitions between adjacent states. Examples of 1-, 2-, 3-ring (task) subspaces are given.

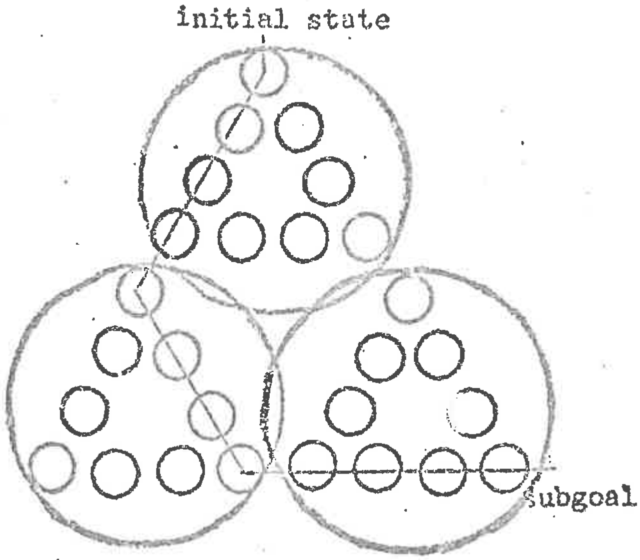

FIGURE 7 a. An'episode' in problem-solving.

The 2-ring subproblem is consistently solved in the minimum number of steps, while the 3-ring subproblem is not. The state-space has been effectively reduced modulo its 2-ring subproblem decomposition.

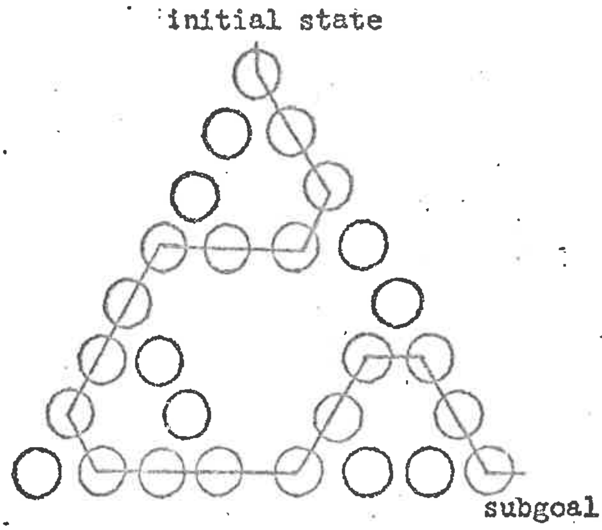

FIGURE 7b. Congruent paths through isomorphic subproblems.

All three paths through the 2-ring subproblems in Figure 6 b are congruent to each other.

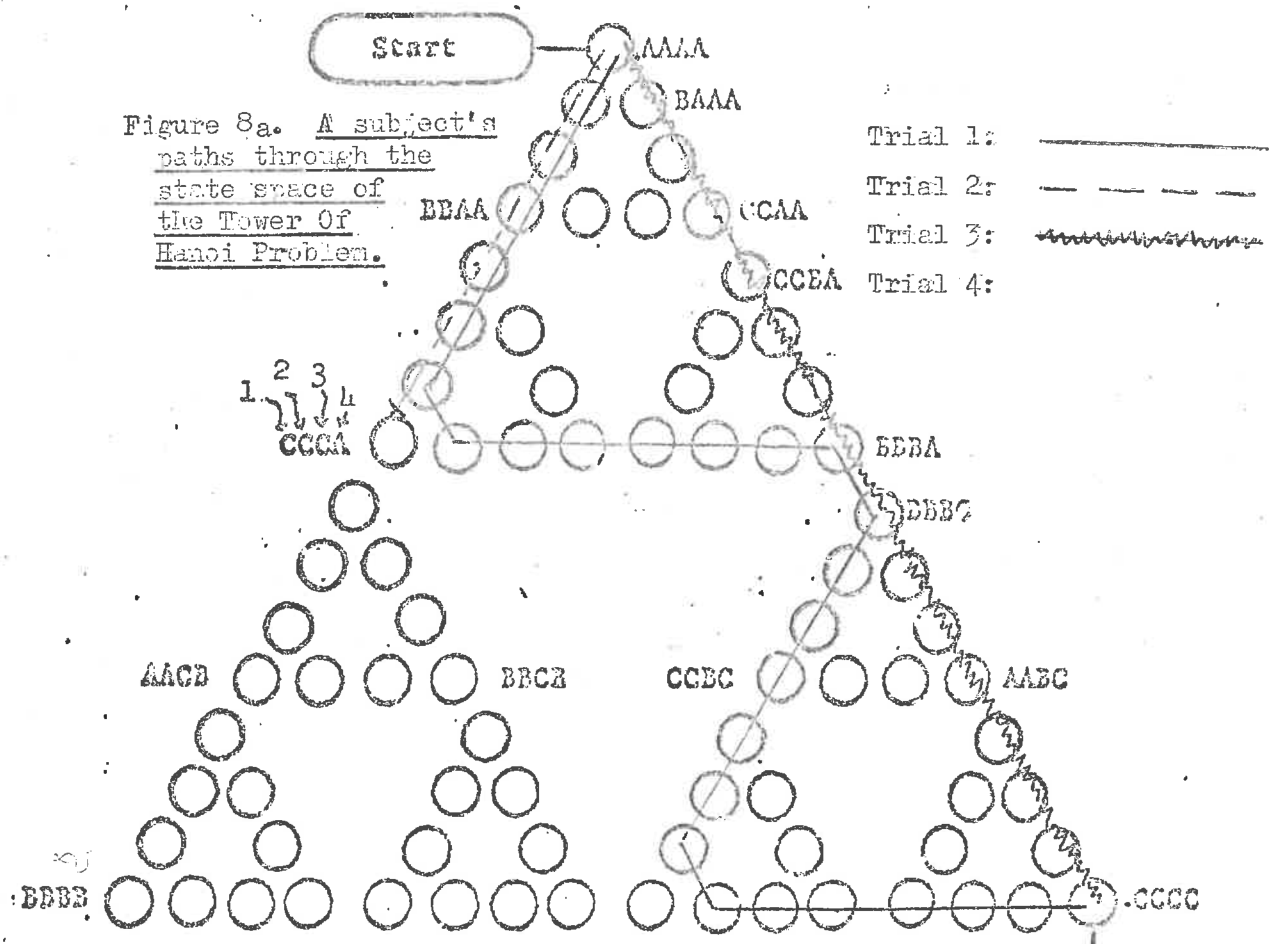

Figure 8a. A subject's paths through the state space of the Tower Of Hanoi Problem.

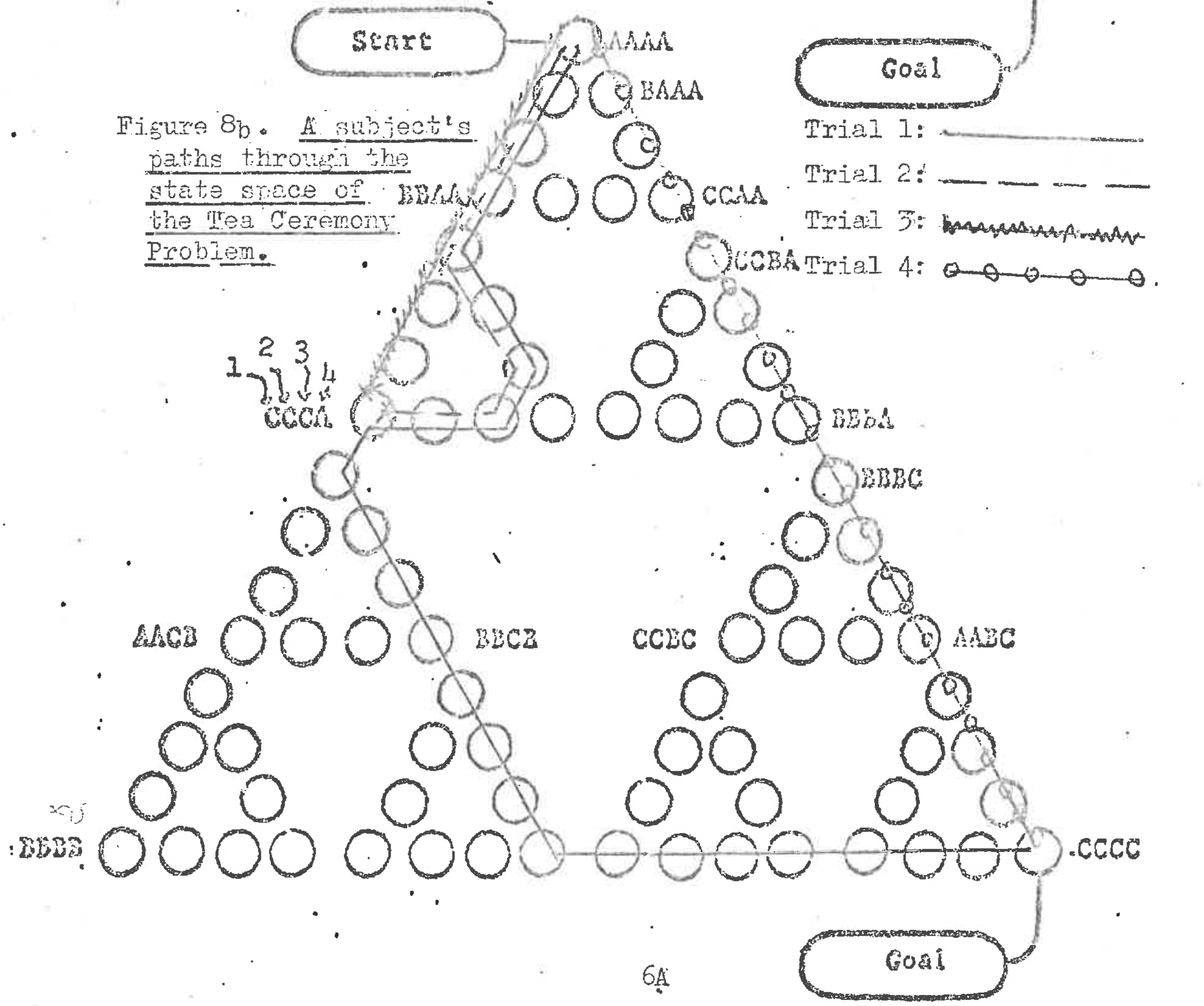

Figure 8b. A subject's paths through the state space of the Tea Ceremony Problem.