1968

The logic theory machine as a theory of human problem-solving

John P. Coyne
Lehigh University

THE LOGIC THEORY MACHINE
AS A THEORY OF HUMAN PROBLEM-SOLVING

by

John P. Coyne

A Thesis
Presented to the Graduate Faculty
of Lehigh University
in Candidacy for the Degree of
Master of Science
in
Information Science

Lehigh University
1968

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

9 September 1968
(date)

Thomas M. Haynes
Professor in Charge

Daniel J. Hillman
Head of the Department

Acknowledgments

I would like to express my gratitude to Dr. Haynes for his many helpful suggestions concerning this work. Also I would like to thank Darien Gardner for making available a copy of his recent work on heuristics and problem-solving which was particularly useful in organizing the material for the first section of this paper.

Work on this thesis was supported, in part, by a National Defense Graduate Fellowship.

TABLE OF CONTENTS

ABSTRACT

In 1956 A. Newell and H. A. Simon (with the aid of J. C. Shaw) published the first paper on the Logic Theory Machine (L.T.). In effect, L.T. was a computer program that proved theorems in propositional logic. Believing that L.T., in solving problems, exhibited many characteristics of human problem-solving, Newell, Simon and Shaw (N.S.S.) began to develop a theory of human problem-solving based on L.T. This paper purports to provide an exposition, critique and evaluation of N.S.S.' work in this area.

Since L.T.'s ability to prove theorems in propositional logic is highly dependent upon the use of heuristics, the first section of this paper provides a definition of "heuristic" and a discussion of the various types of heuristics and their relationship to problem-solving. The second section is devoted to a functional description of L.T. In the third section a presentation is given of N.S.S.' attempt to develop a theory of human problem-solving based on L.T. The fourth section provides an enumeration of the logical, psychological and methodological problems connected with this work of N.S.S. By comparing this work of N.S.S. with certain current works in the psychology of human thought and examining N.S.S.' general approach with the most common ones in the philosophy of the social sciences, the fifth section provides a framework for evaluating the novelty and worth of their approach to human problem-solving. A brief general characterization of the contribution of N.S.S. concludes the paper.

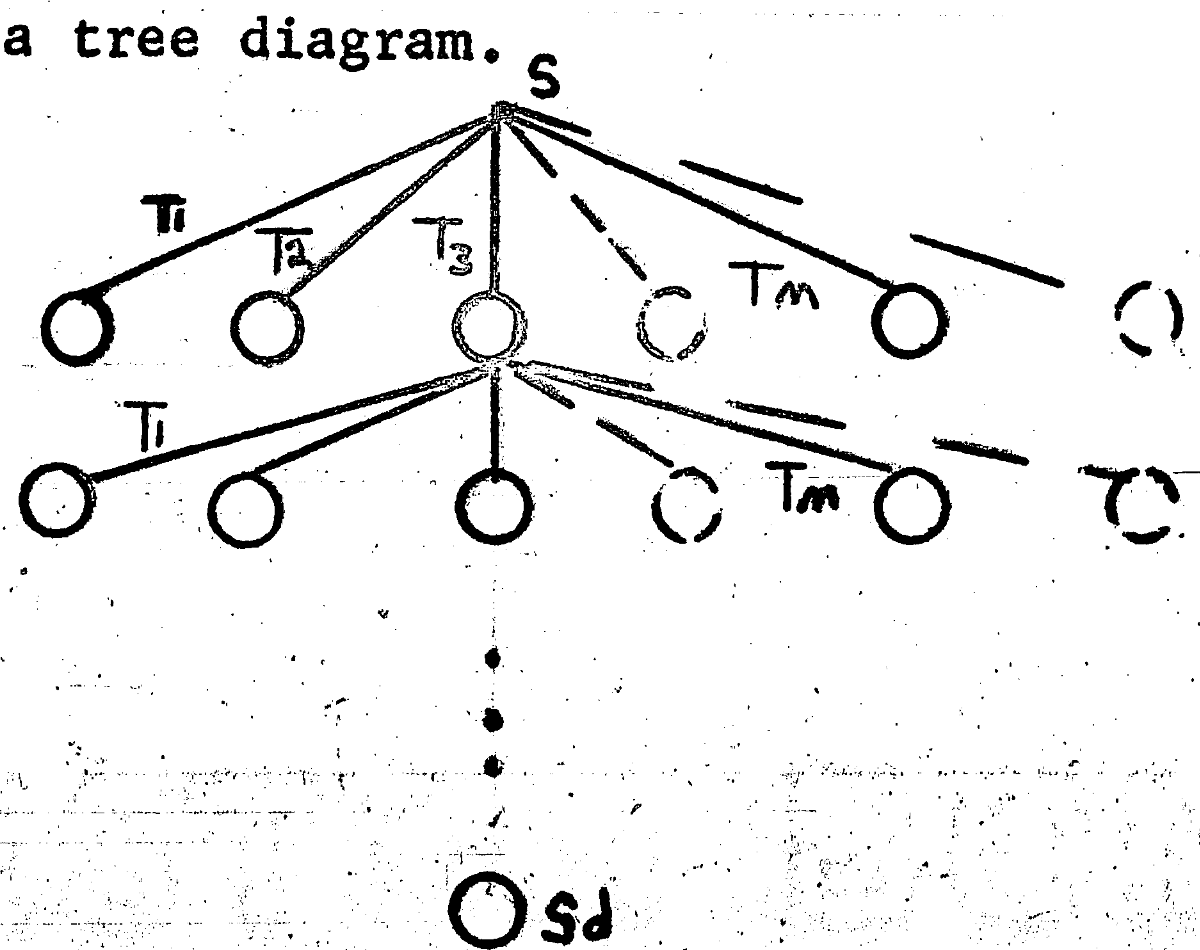
I Heuristics, Problem-Solving and L.T.

The first step in an attempt to clarify the relationship of heuristics to problem-solving is to provide some comprehension of what is a heuristic. A. Newell provides the following definition: (1)"A process that may solve a given problem but offers no guarantee of doing so is called a heuristic for that problem." Although there exists many other definitions,[2] this one will suffice for the present purposes. A heuristic, then, provides one with a hunch. It is a kind of guessing. It is intended to be helpful in the solution of a problem although its fallibility may produce much effort with no success. The most obvious question now is why in the solution of a problem one would want to employ a heuristic which is fallible to an algorithmic process which always produces the right answer, i.e., the problem solution. There are many good reasons, two of which are as follows. First, in many problem-solving areas no algorithms have yet been found. In fact, in certain areas, we know that no algorithm will ever be found. A familiar example of the latter is the first order predicate calculus which is undecidable,[3] that is, no algorithm exist by which for any first order well formed formula (wff) one can decide whether or not that wff is a theorem. A second reason for the selection of heuristics over algorithms is the fact that there exist problem areas wherein an algorithm will always give one a solution but very often the solution is obtained at the cost of a great expenditure of effort. An example of this is the algorithm developed by Newell, Simon and Shaw (N.S.S.) to prove theorems in propositional logic.[4] They estimate that of the sixty-odd theorems that appear in chapter two of Principia Mathematica[5] about six would be included in the first 1,000 proofs generated by the algorithm, but that about a hundred million more proofs would have to be generated to obtain all the theorems in the chapter. By employing heuristic devices L.T.'s performance of the same task is much more efficient.

Although the above examples are provided to illustrate certain advantages of heuristics over algorithms in problem-solving, it is not intended to generate the impression that in constructing problem-solving programs one selects either heuristics, or algorithms, but not both. Indeed, the movement today seems to be toward the development of problem-solving programs that incorporate both heuristics and algorithms in a complementary fashion.[6][7]

Having provided at least an intuitive definition of "heuristic", it is now possible to be more specific and identify four general types of heuristics used in problem-solving.

It is common knowledge that just about every well-defined problem can be viewed as a problem of determining a sequence of transformations leading from a given situation to a desired situation. All possible sequences of applying transformations to the initial situation (S) in an attempt to reach the desired situation (Sd) can be represented in a tree diagram.



In most problem-solving situations, the number of "permissible" transformations is usually finite, and moreover, for any given node, it is usually not the case that all transformations are defined on it. However, the above diagram is an attempt to show the problem tree for the most general situation.

Calling each transformation a method, and calling the initial problem situation and each problem situation that is generated from the initial problem situation by some sequence of methods a "node", the four general types of heuristic generally used in problem-solving are (1) method heuristics, (2) node heuristics, (3) semantic model heuristics, and (4) analogous model heuristics.

(1) A method heuristic is a device that selects certain methods to be applied to a node already chosen. If one had reached the n-th step in an attempted proof that a certain wff in logic is a theorem, then in this context, a heuristic method would be any device that would specify that only certain rules of inference should be tried in attempting to prove the desired theorem from the wff of the n-th step.

(2) A node heuristic is any device that selects from all the nodes generated so far a particular node to work on next. Again, consider the case of trying to prove a theorem in logic. If one has taken a particular axiom and has applied n rules of inference to the given axiom, the result would be n logic expressions each with the property that it can be derived from the given axiom by one application of only one rule of inference. In this context, a node heuristic would be any device that would specify which of the n

logic expression to work on next in an attempt to prove the desired theorem.

(3) H. A. Simon has remarked that:(8) "One theory can have exactly the same logical context as another but be infinitely more valuable than the other if it is stated in such a way as to be easily manipulated, so that its logical context is actually (psychologically) available to the inquirer." When a given formal system is provided with a semantic model, the formal system and its semantic model are identical in logical context, that is, what is accessible by the rules of inference in the formal system is equally (logically) accessible in the semantic model by the corresponding rules. However, an individual may find it psychologically easier to work with the semantic model since it provides "meaning" to the symbols involved. A good example of the use of a semantic model would be in the proof of a wff of a Boolean algebra which can be interpreted as the propositional calculus. Thus, the propositional calculus could serve as a semantic model heuristic. Proof of a wff in the Boolean algebra could be more easily obtained by first solving the problem in the interpreted system, i.e., propositional logic. Since there exists a 1-1 correspondence between a formal system and any semantic interpretation of it, the use of a mapping is all that is needed to obtain the proof in the formal system given a proof in the semantic model.

(4) Analogous model heuristic. An analogous model differs from a semantic model in that there is no 1-1 correspondence between the elements and operations of the problem situation and the elements

and operations of the analogous model. Basically, the use of an analogous model is an attempt to:

(a) transform the original problem into a problem that is easier to solve,

(b) solve the easier problem,

(c) use the sequence of steps in the solution of the easier problem as a guide to a solution of the harder problem.

In the use of an analogous model, there is no guarantee that the steps used in the solution of the easier problem will be useful in determining the steps of the solution of the real problem. As an example of the use of an analogous model in problem-solving consider the following. Suppose one is trying to solve a problem involving a three-dimensional figure. One example of an analogous model heuristic would be the attempt to take a two-dimensional projection of that figure, solve the problem in two-dimensions, and then use the steps in this solution as a guide to the steps to be taken in the solution of the three-dimensional problem.

Most heuristic programs will contain some combination of these four types of heuristics. L.T. uses two of these four types (node selection and method selection heuristics) in proving theorems in propositional logic.

Generally, a heuristic program is characterized by:[9]

(1) division of problem into subproblem,

(2) use of heuristics,

(3) recursiveness,

(4) fallibility.

II A Functional Description of L.T.

In this section, a reasonably thorough functional description of the Logic Theory Machine (L.T.) is given. This description is primarily leased on "The Logic Theory Machine" by N.S.S.(10) although other descriptions exist.(11)(12) From the time of its inception in 1956, several variations of this program have existed. The present discussion confines itself to the original program design. In later sections, variants of L.T. will be noted only when necessary.

L.T. is a complex information processing system capable of proving theorems in propositional logic.(13) Basically, an information processing system consists of a set of memories and a set of information processes. The memories serve as the inputs and outputs for the information processes. A memory is a place that holds information over time in the form of symbols. These symbols are said to function as information entirely by virtue of their capacity to make the information processes act differentially. Viewing them mathematically, the information processes are functions from the input memories and their contents to the symbols in the output memories. The set of elementary information processes is defined explicitly, and through these definitions all relevant characteristics of symbols and memories are specified.

In the above paragraph, L.T. was described as a <u>complex</u> information processing system. Now, by "complex" is meant:

a) there is a large number of different kinds of processes, all of which are important, although not necessarily essential, to the performance of the total system;

b) the uses of the processes are not fixed and invariable, but are highly contingent upon the outcomes of previous processes and on information received from the environment;

c) the same processes are used in many different contexts to accomplish similar functions towards different ends, and this often results in organizations of processes that are hierarchial, iterative, and recursive in nature.

This preliminary notion of a complex information processing system should become considerably clearer from the following description of how the Logic Theory Machine functions.

The function of L.T., that is the task for which it was designed, is to prove that certain expressions in propositional logic are <u>theorems</u> - that is, they can be derived by application of specified rules of inference from a set of primitive sentences or axioms. The language used is the common one for propositional logic. The <u>variables</u> (atomic sentences) are represented by p, q, r, A, B, etc. The connectives used are: -(not), v(or), $\Rightarrow$ (implies). The connectives are used to combine the variables into <u>expressions</u> (molecular sentences) e.g., (-p) $\Rightarrow$ (q v -p). The two connectives, - and v, are taken as primitives. The third connective, $\Rightarrow$, is defined in terms of - and v, i.e., p $\Rightarrow$ q =df - p v q (def. 1.01). The five axioms that are postulated to be true are:

1.2 (p v q) $\Rightarrow$ p

1.3 p $\Rightarrow$ (q v p)

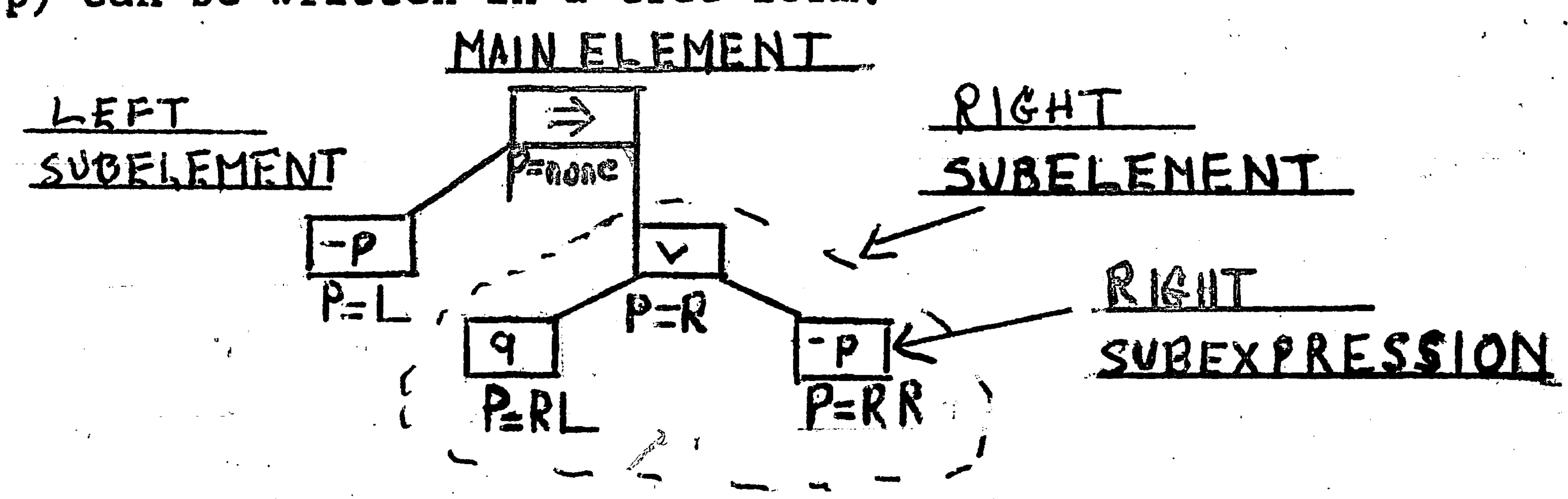
1.4 (p v q) $\Rightarrow$ (q v p)

1.5 (p v (q v r)) $\Rightarrow$ q v (p v r)

1.6 (p = q) $\Rightarrow$ ((r v p) $\Rightarrow$ (r v q))

Each of these axioms is stored on a list in the theorem memory, T, with all its variables marked free, that is,capable of substitution. Now, given an expression to be proven, L.T. attempts of proof by use of three main methods substitution, detachment, chaining.

Before explaining each of these methods, it is best to point out that for any expression L.T. is able to assign to it a unique triple of numbers (K,J,H). This unique triple is determined as follows. Consider the fact that any expression (e.g. -p ⇒ (q v -p) can be written in a tree form:



Now K = the number of levels in the expression. The number of levels corresponds to one plus the maximum number of letters in P for any element in the expression.

J = the number of distinct variables in the expression, ignoring negation signs.

H = the number of variable places in an expression.

Thus, for (-p) ⇒ (q v -p) (K, J, H) = (3,2,3)

Having noted this, attention is now turned to a description of how each of the three methods (substitution, detachment, chaining) is used in proving theorems from the given axioms.

The method of substitution works as follows. Given an expression to be proven, a search is made of the axiom list for an axiom that is "similar" (a notion to be soon defined) to the given expression. When one is found, an attempt is made to "match" (to be defined) it with the expression to be proven. If the match is successful, the expression is proved; if the list of axioms is exhausted without producing a match, the method has failed.

The above notion of "similarity" is now defined. It has already been mentioned how it is possible to compute a unique triple of numbers (K,J,H) for any given expression. Now consider the expression $(p \Rightarrow -p) \Rightarrow -p$. It should be clear that for this expression, (K,J,H) = (3,1,3). Now for this expression (call it expression 2.01), the fact that its unique triple is (3,1,3) is written as: D(2.01) = (3,1,3). In a very similar fashion, one is able to write descriptions for the various subexpressions contained in 2.01--in particular, the sub-expression to the right and to the left of the main connective. Clearly, for each such sub-expressions, DL(2.01) = (2,1,2) and DR(2.01) = (1,1,1) where DR and DL indicate the unique triple (K,J,H) for the left and right sub-expressions (i.e. $(p \Rightarrow -p)$ and $-p$ respectively) of the given expression, 2.01.

Now, two expression, x and y are defined to be similar if they have identical left and right descriptions, that is, if DL(x)=DL(y) and DR(x)=DR(y).

So, two expression, x and y are defined to be similar if they have identical left and right descriptions, that is, if DL(x)=DL(y) and DR(x)=DR(y). The routine for determining whether two theorems are similar, consists of two segments: a description segment and a comparison of descriptions. The description segment is made up of four descriptions routine, one each to compute DL(x), DL(y), DR(x), DR(y). The comparison segment is made up of two compare description routines, one of which compares DL(x) with DL(y) and DR(x) with DR(y). So, given an expression to be proved, L.T. will first compute DR and DL of that expression. Then, L.T. will scan the axiom list to find an axiom that is "similar" to the expression to be proved. Recalling expression 2.01, that is (p⇒ -p)⇒ -p, let us assume that this expression is given to L.T. to be proven from L.T.'s axiom list. It has already been noted that DL(2.01)= (2,1,2) and DR=(1,1,1). The only axiom that is "similar" to this expression is axiom 1.2, that is, (p vp) ⇒ p. Now that L.T. has found an axiom similar to the expression to be proved, L.T. attempts to match the two, that is, L.T. carries out a point-by-point comparison between 2.01, the expression to be proved, and 1.2, the axiom that is similar. In matching the two, one begins with the main connectives, and works systematically down the tree of the logic expressions - always as far as possible to the left. In the present case the order in which one would match is: main connective (P=none), connective of left sub-expression (P=L), left variable of sub-expression (P=LL), right variable of sub-expression (P=LR), and right sub-expression (P=R). To get a clearer understanding of this let us consider the actual steps by which expression 2.01 is matched to axiom 1.2. The matching routine is carried out as follows:

2.01 (p⇒ -p)⇒ -p

1.2 (A v A) ⇒ A

(A is used instead of p in 1.2 in order to indicate that its variable is free).

a. The main connectives agree: both are ⇒

b. Proceeding downward to the left, the connective is ⇒ in 2.01, but v in 1.2. To change the v to ⇒ it is necessary (by def.1.01) to have a - before the left-hand A in 1.2. This can be obtained by making the substitution of -B for A in 1.2. Having carried out this substitution, and having then replaced (-B V -B) with (B⇒ -B), one has the following

2.01 (p⇒ -p)⇒ -p

1.2 (B⇒ -B)⇒ -B

c. Proceeding again to the left, one finds B in 1.2 but p in 2.01. Substituting p for B in 1.2, one finds that after recursion through the remaining two elements that one has a complete match.

Thus, a proof has been discovered of 2.01. Notice that this method of substitution may be viewed as information process that is composed of a considerable number of more elementary information processes arranged to operate in a highly conditional sequence.

Consideration will now be given to the method of detachment -- the second of three methods by which L.T. proves theorems. The principle that underlies this method is simply this: Suppose L.T. is asked to prove that expression A is a theorem; and assume that there are in the theorem memory two theorems, B and B⇒ A. By application of the rule of detachment to B and B⇒ A, A is derivable immediately. In its actual use of this method, L.T. employs a more generalized procedure by combining matching (substitution and replacement) with detachment. Assume that the theorem memory contains B" and B'⇒ A'; that A is obtainable from A' by matching; and that B' is obtainable from B" by matching. Then, it is possible to construct a proof of A as follows:

1) By matching with B", B' is a theorem;

2) Since B'$\Rightarrow$A' is also a theorem, it follows by detachment that A' is a theorem;

3) By matching with A', A is a theorem.

Now with this in mind, the method of proof by detachment can be more fully explained as follows. Let A be the expression to be proved. First, L.T. searches the axiom list for theorems whose right sides are similar to the whole expression A. If such an axiom is found, L.T. then tries to match the right side of this axiom (call it T) to A. If this match is successful, then L.T. attempts to prove that the left side of axiom T is a theorem by the method of substitution. If this is successful, then A has been proven since if the left side of T is a theorem, then the right side of T is a theorem by the method of detachment. But A can be obtained from the right side of T by substitution, thus, A is a theorem. (Note: of course, a check is made to make sure that T has $\Rightarrow$ as a connective).

(It should be noted that the L.T. also employs, if the above method fails, a different criteria of similarity which enables a proof by detachment by use of contraction. However, this routine will not be discussed here since its importance is minor).

Consideration will now be given to the third method of proof that L.T. employs -- that of "chaining." The chaining method may be briefly stated as follows. Given an expression to be proved, that is of the form A$\Rightarrow$ C, an attempt is made to prove that two other expressions, A$\Rightarrow$ B and B$\Rightarrow$ C are theorems. If B$\Rightarrow$ C and A$\Rightarrow$ B can be proved, then by the transitivity of the syllogism, A$\Rightarrow$ C is proved.

To be more specific, given an expression (of the form A=>C) that is to be proven, L.T. does the following. It searches the axiom list for a theorem T (with => for a connective: whose left side is similar to A. An attempt is then made to match the left side of T with A. If this is successful, then L.T. has proven a theorem of the form A=> B, for T, as modified by matching, is of this form. A check is then made to see if it is possible to match B to C. If this is successful, the theorem is proved. If this fails, however, L.T. constructs the expression B=>C and attempts to prove this expression by substitution. If this succeeds, then, L.T. has established the chain A=> B, B=>C and it then concludes that A=> C has been proven according to the transitivity of the syllogism. Now, this procedure just described is known as "chaining forward." L.T. is also capable of "chaining backwards", that is, in order to prove A=>C, L.T. will search for a theorem of the form B=>C, and then it will try to prove A=>B by substitution.

Having described the three methods by which L.T. attempts to prove theorems in propositional logic, attention is now turned to an explanation of the sequence by which these methods are used. Such a sequence or order is determined by what is known as the "executive" routine of the computer program. When a problem is posed to L.T., that is, an expression is presented for which a proof is sought, successive attempts are made to prove the expression by the methods of substitution, detachment and chaining, respectively. To explain what happens if the three methods are unsuccessful, it is necessary to consider the following. It has been noted that both detachment and chaining are two-step methods. Suppose a proof is sought for an expression A. In use of the detachment method, an attempt is then made to find a theorem B= A, and if this is successful, an attempt is then made to prove B. Let us designate the task of proving B a <u>subsidiary problem</u>. Similarly, when an attempt is made to prove A= B by chaining, a theorem of the form A⇒ C is sought. If A⇒ C is successfully proven, an attempt is then made to prove C⇒ B. The task of proving C= B is also a <u>subsidiary problem</u>. Now, within both the detachment and chaining method, only the method of substitution is applied to the subsidiary problem. Now if the three main methods of proof have failed, L.T. then selects the "simplest" subsidiary problem ("simplest" means having the least number of levels, i.e., the smallest K value) and attempts to prove it by means of detachment and chaining. This sequence is repeated until some subsidiary problem is solved or until the number of steps involved exceeds a specified "stop" limit. In the latter case, L.T. reports that it cannot solve the problem.

From the above description, it is clear that L.T. exhibits the general features of a heuristic problem-solving program described earlier. Briefly, those general features are:

1) division of problem into subproblems,

2) use of heuristics, viz.: a) method selection,
   b) node selection,
   c) semantic model,
   d) analogous model,

3) recursiveness,

4) fallibility.

The first of these four is exhibited in many ways, the most notable being the production of what has been described as <u>subsidiary problems</u>.

With respect to heuristics, L.T. makes use of a method selection heuristic and a node selection heuristic.

For a given axiom and a given wff to be proven, the notion of "similarity" is used to limit the methods of proofs that will be applied to that axiom in attempting a proof. Thus "similarity" provides a method heuristic for L.T.

In selecting which subsidiary problem is to be worked on, L.T. employs a notion of "simplicity" (the smallest K value). So "simplicity" is a node heuristic for L.T.

Recursiveness is evident in L.T. in many ways. Certainly, the many iterations of attempting proofs by substitution is a good example.

L.T. is certainly fallible, that is, it does make mistakes. Consider the fact that the wff $p \rightarrow (p \vee p)$ can be simply proven by substitution of p for q in the axiom $p \rightarrow (q \vee p)$. However, this axiom would not be considered for proof by substitution since it fails to satisfy the notion of "similarity" since their J values (number of distinct variables) are different.

III L.T. and a Theory of Human Problem-Solving

Having described in reasonable detail the operation of L.T., discussion is now turned toward its capacity to provide a basis for (or elements of) a theory of human problem-solving.

The proposed theory of human problem-solving is based on a program: a specification of what the organism will do under varying environmental circumstances in terms of certain elementary information processes that it is capable of performing. Newell, Simon and Shaw state that if one considers the organism to consist of effectors, receptors and a control system for joining these, then this theory is mostly a theory of the control system. The theory postulates a control system consisting of a number of memories which contain symbolized information and are interconnected by various ordering relations, a number of primitive information processes which operate on the information in the memories, and a perfectly definite set of rules for combining these processes into whole programs of processing. Thus, an explanation of an observed behavior of the organism is provided by a program of primitive information processes that generates this behavior.

The program of L.T. can be used as a theory (in the sense of a predictor of behavior) in two distinct senses according to Newell, Simon and Shaw.(14) First, it is able to make many precise predictions that can be tested in detail regarding the area of behavior it is designed to handle. Secondly, it will provide qualitative characteristics of human problem-solving which can be compared with characteristics already described in psychological literature. For L.T., Newell, Simon and Shaw limit their comments to this second type of validation of their theory since "all of the available data on the psychology of human problem-solving are of this qualitative kind..."

A summary will now be given of the resemblance of certain characteristics of L.T.'s program with aspects of the human problem-solving process as it has been described in psychological literature. These resemblances, according to Newell, Simon and Shaw may be summarized under the following headings: set, insight, concept formation, and structure of the problem-subproblem hierarchy. In addition, they think that L.T. illustrates some, though not all, of the forms of human learning.(15)

As Newell, Simon and Shaw point out their term "set" has sometimes been defined as "a readiness to make a specified response to a specified stimulus," and it (the term) covers a wide variety of psychological phenomena. The behavior of L.T. exhibits "set" in several aspects. Moreover, these several evidences of set correspond to quite different underlying processes. First, consider the fact that after the program has been loaded in the hardware (computer) of L.T. and the axioms stored in its memory, before L.T. will attempt to prove the first problem it is given, it will first go through the list of axioms and compute, <u>for each axiom</u>, the unique triple (K,J,H) for the whole expression and the unique triples for each of the sub-expressions. As Newell, Simon and Shaw point out, this computation process and the interval of time required for it, represent (functionally and phenomenologically) a <u>preparatory set</u>

in the sense it is used in reaction-time experiments. Second, consider the fact that when L.T. is attempting a particular subproblem it tries to solve it by the substitution method and then only if that fails does it employ the detachment method and then chaining. Now, when it searches for theorems suitable for the substitution method, it will not take any notice of theorems that might later be suitable for the detachment (different similarity tests being applied in the two cases). L.T. pays exclusive attention to possible candidates for substitution until the theorem list has been exhausted - only then will L.T. employ the detachment method. Newell, Simon and Shaw believe that this behavior represents a directional set.

With respect to the notion of "insight", Newell, Simon and Shaw first note that in psychological literature "insight" has two principal connotations: "suddenness" of discovery, and grasp of the "structure" of the problem as evidenced by absence of overt trial-and-error. Newell, Simon and Shaw believe that L.T.'s program partially resolves the "insight" vs. "trial-and-error" debate. L.T. does attempt to prove theorems by trial-and-error procedures; but trial-and-error attempts take place in a limited space of all possible solutions. For example, the use of "similarity" tests greatly reduces the number of possible solutions that will be attempted. Thus, trial-and-error is not a totally arbitrary technique. The sequence of possible solutions that L.T. attempts does, in fact, depend upon problem "structure."

With respect to concept formation, Newell, Simon and Shaw point out that L.T. is primarily a performance program. Nevertheless, L.T.'s program does supply a clear example of concept use in problem solving. This is found in the routine for describing theorems and searching for theorems "similar" to the problem expression or some part of it in order to attempt substitutions, detachments, or chainings. As Newell, Simon and Shaw state, "All theorems having the same description exemplify a common concept."

Concerning problem-subproblem hierarchy, L.T. employs two types of hierarchies. In attempting to solve a problem, L.T. tries proofs by substitution, then detachment, and then chaining. The second more interesting kind of hierarchy is the generation of new expressions to be proven. It has been noted earlier that both detachment and chaining method do not give proofs directly, but, instead, provide new alternative expressions to prove. A list of such subproblems is kept by L.T. and L.T. is able to apply all its problem-solving methods to them. Moreover, these methods yield yet other subproblems, and thus a large network of problems is developed during the course of proving a given logic expression. The interesting aspect of such a hierarchy is that it is flexible, it grows in response to the problem-solving process itself.

Newell, Simon and Shaw believe that the problem-subproblem hierarchy in L.T.'s program is quite comparable with the hierarchies that have been observed in studies of human problem-solving.

With respect to learning, N.S.S. believe L.T. possesses a number of important learning processes that exhibit some forms of human learning.

N.S.S. define learning as any more or less lasting change in the response of the system to successive presentations of the same stimulus. By this definition, L.T. learns in the following ways.

(1) When L.T. has proven a theorem, it stores this theorem in its memory. Henceforth, the theorem is available for the proof of subsequent theorems. Therefore, whether L.T. is able to prove theorems depends, in general, on what theorems it has previously been asked to prove.

(2) L.T. remembers, during the course of its attempt to prove a theorem, what subproblems it has already tried to solve. If the same problem is presented twice in the course of the attempt at a proof, L.T. will remember and will not try to solve it a second time if it has failed a first.

(3) In one variant, L.T. remembers what theorems have proven useful in the past in conjunction with particular methods and tries these theorems first when applying the method in question. Hence, although its total repertory of methods remains constant, it learns to apply particular methods in particular ways.

N.S.S. believe that the several types of learning now found in L.T. begin to cast light on the pedagogical problems of "what is learned" including the problems of transfer of training. For example, if L.T. simply stored proofs of theorems as it found these, it would be able to prove a theorem a second time very rapidly, but its learning would not transfer at all to new theorems. The storage of theorems has much broader transfer value than the storage of proofs, since, as already noted, the proved theorems may be used as stepping stones to the proofs of new theorems.

IV Problems

We now turn to a consideration of certain questions concerning N.S.S.' approach to human problem-solving. Before this is done, however, let us give a reasonably clear summary of what they are attempting to do in the way of theory formation.

Their proposal is that an individual solving problems be regarded as an information processing system - a notion explained earlier. At their level of theorizing, "an explanation of an observed behavior of the organism is provided by a program of primitive information processes that generate this behavior." Viewed as a theory of behavior, a program is highly specific in that it represents only the behavior of one individual in one set of situations. If either the individual or the class of situations is changed, the program must be changed. N.S.S. hope, however, that important similarities will occur among the programs which represent the behavior of the same individual in different situations, or among those which represent the behavior of different individuals in the same situation. Hopefully, a more general theory of the kind of behavior under study may be developed. Finally, we note that in employing this approach to problem-solving, or some other area, one begins by attempting to identify the processes involved in the particular kind of thinking under study, and when some hypotheses have been formed as to the processes involved, one then attempts to write a program which employs these processes and which simulates the thinking of the human individual. The objective in writing the program is not only to achieve the results which the human thinker achieves, but also to employ the same processes in doing so.

Now, with this in mind, consider again the above definition of explanation -- "an explanation of an observed behavior of the organism is provided by a program of primitive information processes that generate this behavior." This appears to be a particularly bad conception of what constitutes an explanation for the following reasons.

Consider a problem-solving area wherein there are only a finite number of individual problems (say n such problems). A particular individual is asked to solve each of the n problems, and to write on scratch paper (or say aloud) every step he takes. Let us presume that he solves the n problems with no difficulty and we have taken care to collect his scratch paper or recorded his talking while working. Let us presume further that if we present him with these n problems at another time there will be no significant difference (assuming for the moment that the notion of "significant difference" is intuitively clear) between his original solution of a problem and his solution of the same problem at a later time. Recall that an information processing program consisted of:

1. a control system consisting of a number of memories, which contain symbolized information and are interconnected by various ordering relations;
2. a number of primitive information processes which operate on the information in the memories;
3. a perfectly definite set of rules for combining these processes into whole programs of processing.

It is a trivial (though possibly tedious) task to write an information processing program (in fact, a "complex" one as N.S.S.

use the term) such that when one of the above n problems is presented to it, it prints out the steps that our individual under discussion would use in the solution of the problems. In short, for each problem we simply store in memory the exact information that we obtained from the individual's scratch paper and/or recording, and simply write a program to print out this information when the same problem is input to the computer. Now such a program would satisfy N.S.S.'s definition of "explanation." Yet, in what sense have we really explained anything about this individual's problem-solving ability? This program no more "explains" this individual's problem-solving ability than a parrot that has been trained to utter certain words (the problem solution steps) when he hears other words (statement of the problem).

In the discussion above, we mentioned that N.S.S. were attempting to develop theories of problem-solving for a given individual in a given area by constructing programs that were able to generate the same behavior as the individual and whose processes exhibited problem-solving features observed in human problem-solving. Let me now make a three-fold distinction among:

1) mechanisms,
2) processes,
3) features.

These three are related as follows. By mechanisms is meant the actual physical device upon which a given process or combination of processes are <u>realized</u>, that is, the processes are implemented on the mechanisms. In our case, the combination of processes in a certain way is the program whereas the mechanism is the computer. Features are certain aspects or notions <u>exhibited by</u>

a process or combination of processes. With this distinction in mind, let us consider several questions concerning Newell, Simon and Shaw's approach to a theory of human problem-solving.

The first point we wish to consider is how one begins to develop a theory for an individual in a given problem-solving area. At first, the method seems quite simple and direct. At least the data gathering is simple and direct: the data is obtained from a human subject who is asked to use scratch paper as he works a problem, or to think aloud. In the next step, one attempts to identify the processes that the given individual employs in solving the problem in the given area. At this step certain questions arise. The first question is <u>how does one</u>, given a certain amount of data, "identify" the processes that the individual uses. Now, one cannot directly observe a process or processes, but only the outputs of these processes. So there seem to be either of two alternatives. First, the "identification" of these processes can be an <u>ad hoc</u> affair. By this is meant that the observer somehow intuitively perceives the processes. The lack of control and scientific rigor for such an approach is obvious. The other alternative is that the observer uses a specific technique or measure to decide for a given amount of data just what processes are being employed by the individual in problem-solving. Of course, this approach causes one to question just what such a technique or measure would look like. Further, the question arises as to whether or not such a measure or technique is unique or if there are other similar measures or techniques. If the former, then what guarantee do we have of its uniqueness? If the latter, by what criterion or criteria do we decide which one to employ?

In all probability, N.S.S. would assert that in their work, as in other areas of science, one gathers certain data on a given range of phenomena, and based on this data one begins to conjecture hypotheses and then tests these hypotheses.(16) In the area of problem-solving, one would observe the data (scratch work or recording) of a given individual solving a certain number of problems, and then the observer would speculate as to the processes involved, attempt to write a program which employs these processes, and then attempt to test the hypotheses. For the moment, let us not consider the question of verification but instead a possible built-in danger concerning N.S.S. approach. In order to understand this, let us first briefly examine some remarks on how visual experiences become organized -- how seeing is possible. Consider figure one in the context of figure two.(17)

fig. 1 fig. 2 fig. 3

Some people could not visualize figure one as the head of an antelope. It is the context that gives one the clue. Consider figure three. Some people might claim that figure two has no similarity to figure three, although figures two and three are congruent. With respect to figure four, Wittenstein writes:(18)

> You could imagine this appearing in several places in a text-book. In the relevant text something different is in question every time: here a glass cube, there an inverted open box, there a wire frame of that shape, there three boards forming a solid angle. Each time the text supplies the interpretation of the illustration. But we can also see the illustration now as one thing, now as another. So we interpret it, and see it as we interpret it. (Italics mine)

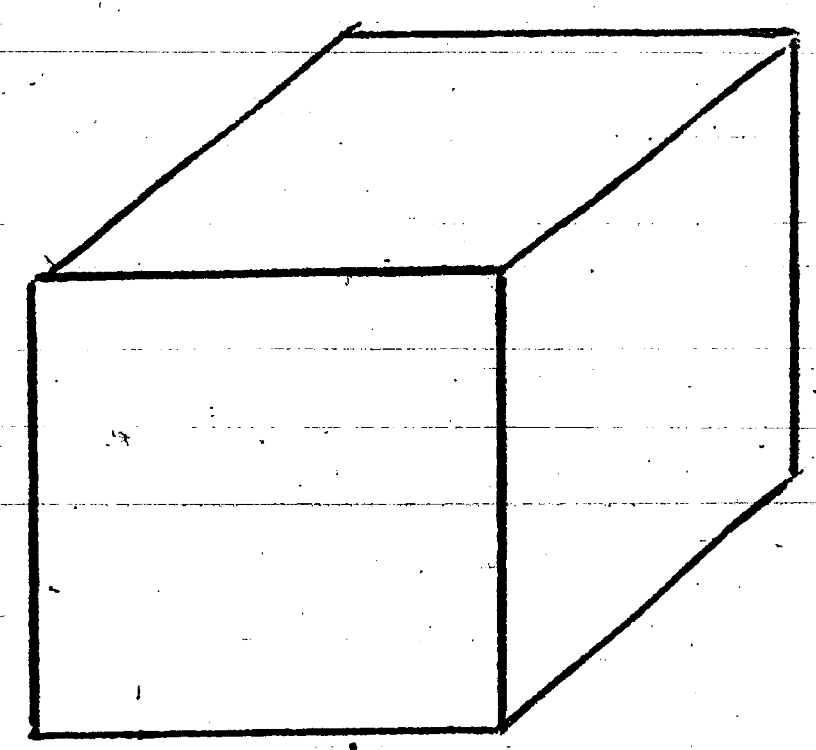


fig. 4

However, the point is that the context need not be set out explicitly. Often it is "built into" thinking, imagining and picturing. In a sense we are set to appreciate the visual aspect of things in certain ways. Elements in our experience do not cluster at random. Consider yet another example -- figure five.(19)

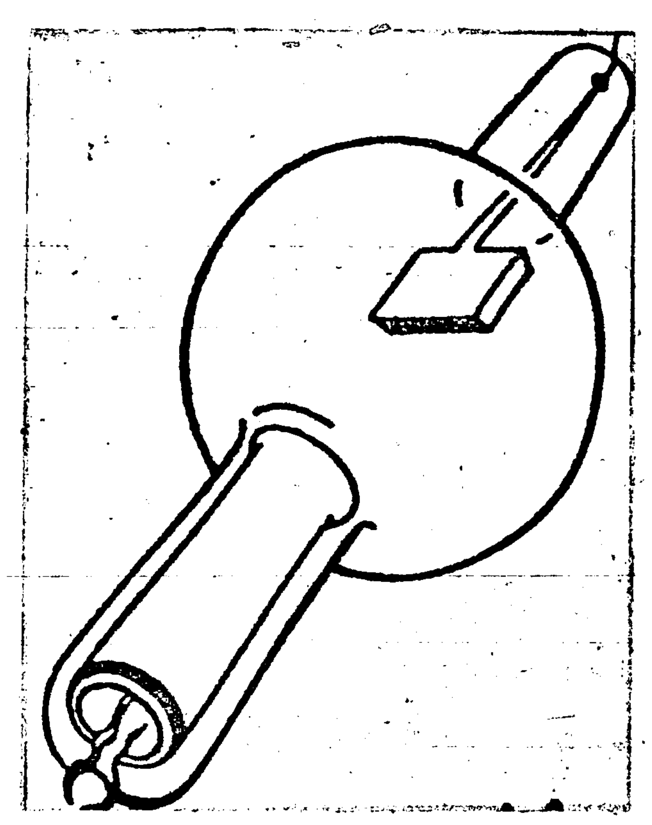


fig. 5

When looking at figure five, a trained physicist would see an X-ray tube viewed from the cathode. Would an Eskimo baby see the same thing when looking ~~at~~ an X-ray tube as Sir Lawrence Bragg would see? Yes, in the sense that they are visually aware of the same object (i.e., identical photons are reflected from the object; these travel through the atmosphere. The two people have normal vision; hence these photons pass through the cornea, aqueous humous, iris lens, and vitreous body of their eyes in the same way. Finally their retinas are affected. Similar electro-chemical changes occur in their selenium cells. The same configuration is etched on Bragg's retina as on the Eskimo baby's). However, although they are both visually aware of the same object, the ways in which they are visually aware are considerably different.(20)

The point of all this is to illustrate that there is a sense, an important sense, in which seeing, or observation, is a "theory-laden" undertaking. Observation of x is shaped by prior knowledge of x. In other words, interpretation of x is not something that is done ex post facto of observing x. Interpretation is often built-in. The trained musician who hears that an oboe is out of tune does not hear the tones and interpret them as being out of tune, but will simply hear the oboe to be out of tune.(21)

What has all of this to do with N.S.S.? Well, it appears that N.S.S. have a definite built-in interpretation that would affect their observation. Since their whole approach is based on the notion of a program, any observation of problem-solving activity will only involve "seeing" that portion of the data for which a process can be constructed that will give this and only this data as output. In other words, N.S.S. would "see" only that subset of the data for which a program can be written which will produce that and only that data. One may ask for any set of data if a program can be written to yield that and only that data as output. Let us consider a more specific instance.

Let I be the "set of integers, i.e., 0, ± 1, ± 2, ..." Now let us consider some sequence of integers, that is, some set of integers in some order. We now ask the question whether or not for any such set of integers we can construct a program (a finite, non-empty ordered list of instructions) that will print out each member of and no other integers. The answer is no.[22] Phrased mathematically, it means that we know that there are certain set of integers that are not recursively enumerable where a set G (of integers) is recursively enumerable if there is a program which prints out each member of M and no other integers. (Note: we have not given a rigorous definition of "program" since it would be too involved. However, the term is being used the same way throughout and does mean what we call "computer program").

Now, to show the possible built-in danger of N.S.S. approach, we have only to show the relationship of the above to their work. The argument will be one from analogy -- hopefully a good analogy. Let us assume for the moment that human problem-solving behavior can be quantified, that is, we could talk about "units" of problem-solving behavior. Let us then consider P to be the set of all such units.

Now why isn't it possible that a person's problem-solving behavior not be recursively enumerable - i.e., you can't write a program to reproduce it. One might object and say that the argument based on our mathematical analogy is a bad one for the following reason. The only sets that are non-recursively enumerable (and there can be no program to reproduce the sets) are infinite sets whereas even if we consider the problem-solving activity of the individual's whole life it is still finite. This objection is based on a misconception of what a problem-solving theory or any theory purports to do. A problem-solving theory does not tell you what the lifetime problem-solving activity will be for an individual; rather a problem-solving theory is of the form "if......then......" whether or not the "if" condition ever actually occurs. Thus, a problem-solving theory would not be limited to a finite number of instances. Indeed, a problem-solving theory of propositional logic would purport to be able to describe an individual's problem-solving behavior in proving any of the wffs.(23)

$$p \supset (p \vee q)$$

$$p \supset (p \supset (p \vee q))$$

$$p \supset (p \supset (p \supset (p \vee q)))$$

$$\vdots$$

Clearly, then, a problem-solving theory able to describe problem-solving behavior cannot be limited to a finite number of instances. Thus, we cannot argue that a problem-solving theory will only concern itself with a finite amount of problem-solving behavior and that our mathematical analogy is not a sound one. Indeed, the above analogy may not be sound (though not for the reason just discussed), however, I think its function is still useful. It conveys the point that N.S.S. approach is <u>predicated</u> on the belief that a thorough account of problem-solving behavior can be described by means of a program (which, in actuality, is a formal closed system) and that in their subsequent development of problem-solving theory this premise is not questioned.(24)

Thus, their observation of problem-solving behavior is going to be myopic in the sense that they will only "see" problem-solving behavior that can be so described. Moreover, this becomes a vicious cycle. Because one begins with the belief that problem-solving behavior can be thoroughly described by a program (or set of programs) one only "sees" problem-solving behavior that can be so described. The more successful one is in "seeing" this way, (i.e., the more one "forces" problem-solving data to be described by such a program) the more entrenched this original premise becomes and the less open it is to question. This again distorts our subsequent "seeing" even more. This is the built-in danger embedded in N.S.S.' approach encompasses. Moreover, this problem does not only show itself in the area of "observation of data" but in others, as we shall see.

Another area of N.S.S.' investigation that seems questionable is that of verification. Recall that N.S.S.(25) believe that a theory of problem-solving (say, for a given individual in a given class of situations) is capable of verification (or validation) in two ways:

(1) We can compare the problem-solving features (or qualitative characteristics) that are exhibited by the processes of L.T. with qualitative characteristics (or features) of human problem-solving behavior that have been observed in the literature of psychology.

(2) For a given problem or problem (in the above class of situations) we can compare the computer print-out of the steps it took in problem solution (or attempted solution) with the steps of the given individual recorded on tape or on scratch paper.

Let us now consider the first type of verification. Earlier, we gave an account of the resemblances of certain features of L.T. with aspects of the human problem-solving process(es) as it had been observed in the psychological literature. We grouped these resemblances under the headings: set, insight, concept formation, and structure of the problem-subproblem hierarchy. In addition, we mentioned that N.S.S. believe that L.T. illustrates some, though not all, of the forms of human learning. With respect to this first form of validation or verification, several questions arise. First, let us presume that there exists a list (an accurate one) of the m characteristics of human problem-solving. (We ignore for the moment how this list was compiled). Let us assume that L.T. exhibits k such features (where k is less than m). So, there exists m-k problem-solving features (or qualitative characteristics) that L.T. does not exhibit. What interpretation do we supply to the absence of these m-k features? Does it mean that L.T. is not a valid theory, or does it mean that L.T. is nevertheless valid because it exhibits k of the m features, or does it mean that L.T. is $\frac{k}{m}\%$ valid (whatever that might mean)? In short, given that L.T. exhibits k of the m features of human problem-solving, how do we then compute the extent to which L.T. does or does not provide a valid theory?

The above are all valid questions. However, there are even more fundamental questions that we should ask concerning verification of L.T. in this first sense. This list of features of human problem-solving -- how is it compiled? N.S.S. are apparently working with those features cited most often in the psychological literature on human problem-solving. However, if L.T. (initially at least) is supposed to represent, or provide a theory of problem-solving for a given individual in a given class of situations, then how do we know that all of these n characteristics are relevant? Certainly, it would be a strong assumption to believe that for a given class of situations, every individual will exhibit certain problem-solving features. Similarly, for a given individual, it is a strong assumption to believe that he will always exhibit the same problem-solving features no matter what the class of situations.

Moreover, there is still an even more fundamental difficulty here. Let us assume that for a given individual in a given class of situations, we have, by some means, compiled a list of m problem-solving characteristics (or features) that he exhibits in solving or attempting to solve problems in that class of situations. Let us assume, again, that L.T. exhibits only k such features (where k is less than m). Do we admit then that L.T. does lack these m-k features, or do we argue that these m-k features were not "real" problem-solving characteristics (or features) of this individual at all? In short, will we only "see" (here the problem arises again) as true problem-solving features those features that can be exhibited by some process in L.T.?

Lastly, if we do accept this list of n problem-solving characteristics as being accurate, we must ask that list is or was verified!

In short, N.S.S. proposal that L.T.'s theory can be validated (or verified) by comparing its qualitative problem-solving characteristics (or features) with those observed in the psychological literature appears to be very ill-defined and questionable for the above reasons.

Earlier we noted that for L.T., N.S.S. attempted at least a partial verification of this first type. This partial verification consisted of showing how L.T. exhibited certain problem-solving features found in the psychological literature. The particular problem-solving features that N.S.S. selected for discussion were: set, insight, structure of the problem-subproblem hierarchy, and concept formation. In addition, they maintained that L.T. exhibits some, though not all, aspects of human learning. Before we examine their second type of verification it might be worthwhile to show that in addition to the above noted fact that they fail to supply a clear definition of the general aspects of their first type of verification, they have also failed to present a convincing argument that L.T. exhibits the problem-solving features noted above in any significant way. In particular their argument that L.T. exhibits set, concept formation and aspects of learning appears to be inconvincing in certain respects.

We have noted before that N.S.S. claim that L.T. exhibits "directional set" for the following reasons. When L.T. is attempting to solve a subproblem it tries to solve it by the substitution method and then only if that fails does it employ the detachment method and then chaining. Now when it searches for theorems suitable for the substitution method, it will not take any notice of theorems that might later be suitable for the detachment method. L.T. pays exclusive attention to possible candidates for substitution until the theorem list has been exhausted - only then will L.T. employ the detachment method. N.S.S. believe that this behavior represents a directional set.

The above claim that L.T. exhibits directional set appears objectionable for the following reason. If we accept the above behavior of L.T. as exhibiting directional set then we have trivialized this notion. Consider the fact that any problem-solving 'organism' (human or otherwise) in attempting to solve a problem must search for the solution in a space (often very large) of possible solutions. Now heuristics may be used to limit our search to a certain subset of possible solution. But even after the heuristics have eliminated our consideration of certain possible solutions, we are still faced with the task of examining the remaining possible solutions. Clearly, the possible solutions must be examined in some particular sequence. If we accept the above cited behavior of L.T. as exhibiting directional set, then it follows that any approach by any 'organism' (human or otherwise) to solving problems involves or exhibits directional set. But if this is so, then we cannot regard directional set as being particularly distinctive in any way of human problem-solving. Yet if this is the case, then in what sense does the fact that L.T. exhibits directional set contribute to verifying that L.T. can provide a theory of human problem-solving? The fact that any problem-solving 'organism' will exhibit directional set makes N.S.S. observation of L.T.'s exhibition of it seem trivial if not redundant. If every problem-solving 'organism' will exhibit directional set, then in what way can this fact be of value in attempting to verify that L.T. can provide a theory of human problem-solving? (It appears that similar remarks could be made about N.S.S. claim that L.T. exhibits preparatory set).

With respect to "concept formation" N.S.S. remarks appear particularly puzzling. Consider the following two quotes from the same paper:(26)

> Using as our data the information provided by L.T. as to the methods it tries, the sequence of these methods and the theorems employed, we can ask whether its procedure shows any resemblance to the human problem-solving process as it has been described in psychological literature. We find that there are, indeed, many such resemblances, which we summarize under the following headings: set, insight, concept formation, and the structure of the problem-subproblem hierarchy.

> The current version of L.T. is mainly a performance program, and hence shows no concept formation. (Italics mine)

Whatever their motivation for introducing the phrase "concept formation" since L.T. doesn't exhibit it, they nevertheless, do maintain that their program exhibits the use of concepts. In this same paper they state:

> There is in the program, however, a clearcut example of the use of concepts in problem solving. This is the routine for describing theorems and searching for theorems "similar" to the problem expression or some part of it in order to attempt substitutions, detachments, or chaining. All theorems having the same description exemplify a common concept.

Now the objection we wish to raise is not that L.T. doesn't exhibit the use of concepts but that it doesn't exhibit it for the reasons in the above quote. To have a concept is to have a certain ability. It is the ability to operate with a certain symbol - to know when to apply the symbol to some object or experience, and to know when to refrain from applying this symbol. In order to have a concept one must have some criterion that enables one to determine when the symbol is applicable and when it is not. That is, we must have some way of consistently applying our symbol. (This is basically the point that Wittgenstein uses in arguing against the possibility of a "private language").(27) Now L.T. seems to use the notion of "similarity" in just this way. L.T. has a test (described earlier on page 9) to determine whether or not two logic expressions are similar or not. Thus it seems that L.T. has, or makes use of, a concept of "similarity."

The objection to N.S.S.' claim that L.T. exhibits the use of concepts is due to the fact that N.S.S. claim that all logic expressions that have the same triple (H,J,K) exemplify a common concept. This seems objectionable because it identifies a concept with an extensionally defined class and does not emphasize the necessity of having a criterion or criteria for deciding when a concept does or does not apply to a given object or experience. Moreover, the identification of a concept with an extensionally defined class seems to lead to a trivialization of the notion "concept" for the following reason. Consider an extremely simple Turing Machine.(28) This machine has only two internal states -- call them x and y. Its input is logic expressions - i.e., we feed into the machine only logic expression one at a time. Its output is either the letter a or b. Now its rules of operation is as follows:

> If the machine is in state x and we input a logic expression, the machine will output (assume it writes its output down on paper) the letter "a" and its internal state will change to "y".

> If the machine is in state "y" and we input a logic expression, the machine will output the letter "b" and its internal state will change to "x".

Now assume we begin to feed into this machine every possible logic expression (actually, since the number of such expression is denumerably infinite, we will never finish the job). It is obvious, that if the machine is initially in state "x" ("y") the first logic expression will cause the machine to print out "a" ("b"). The second such expression will cause the machine to print out "b" ("a"). Obviously, if all the odd numbered logic expressions are mapped to "a", all the even numbered ones will be mapped to "b" or vice versa. Now if we accept N.S.S.' remarks it seems reasonable to declare that all the logic expressions that have been mapped to "a" exemplify a common concept and all the logic expressions that have been mapped to "b" exemplify a common concept. Clearly we would not want to admit that this was so since we have no notion at all as to the nature of this "common" concept. Yet if we accept N.S.S.' reason as justification for claiming that L.T. uses concepts, then we should accept our Turing Machine as using concepts. The fact is we wouldn't accept it as using concepts for many reasons. Perhaps the most outstanding is the fact that our machine lacks any criterion for consistent mapping of a given logic expression, that is, when we input the same logic expression a second time it may not be mapped to the same class as it was the first time. This is sufficient reason to deny any claim that the Turing Machine in question exhibits the use of concepts unless we wish to trivialize the notion of "concept" by maintaining that any "organism" that assigns objects to classes (whether or not this assignment is arbitrary and inconsistent) uses concepts. In short, we wish to maintain that L.T. does use concepts. But this claim rests on the fact it uses, or has, a concept of similarity because it has a criterion (a test) to determine whether or not the symbol "similarity" is applicable to two logic expressions, and not on the claim that all logic expressions that have the same (H,J,H) triple exhibit a common concept.

In declaring that L.T. exhibits use of concepts care should be taken to note that L.T.'s use of concepts does differ in part from the use of concepts by the human mind. It differs at least in the sense that when human beings (members of a linguistic community) acquire a concept, this acquisition is not independent of other concepts.(29) In short, one does not come to understand a language by coming to understand its constituent parts: the words, phrases, and sentential forms of which it is composed. For example, one cannot fully understand "I am in pain" without understanding "He is in pain" and vice versa. Thus neither of these sentences can serve as independent stepping stones to the mastery of English. Unless one already knew a great deal of English such as the use of English pronouns, the forms of the verb "to be," the grammar of the phrase "to feel", the full significance of these sentences would be far beyond one's understanding.

The above aspects of concept use in English (or any other language) seems to have no corresponding aspect in L.T. use of concepts.

With respect to "learning", we have noted earlier that N.S.S. believe that L.T. displays some forms of human learning. In particular we noted that:

(1) When L.T. has proven a theorem, it stores this theorem in its memory. Henceforth, the theorem is available for the proof of subsequent theorems;

(2) L.T. remembers, during the course of its attempt to prove a theorem, what subproblems it has already tried to solve. If the same problem is presented twice in the course of an attempt at a proof, L.T. will remember and will not try to solve it a second time if it has failed a first;

(3) In one variant, L.T. remembers what theorems have proved useful in the past in conjunction with particular methods and tries these theorems when applying the method in question.

We wish to show that these three instances of "learning" certainly do not exhaust our general notion of learning. In fact, the above three instances seem to be particularly weak instances of "learning." Let us consider each in turn.

The first type of "learning" appears objectionable because if the theorems (already proven) were erased from memory and L.T. was again presented with these theorems to prove, L.T. would go through exactly the same steps as before. In contrast, we expect significant learning to use genuine generalization and induction in processing its experience.(30)

The second type of "learning" may be questioned on the following grounds. If we are attempting to solve a problem "a", we may eventually attempt to solve "a" by solving some subproblem "a'". If we attempt to solve "a'" and find after n steps in our proof that we are again faced with proving "a'" (i.e., we start with subproblem "a'" and after much work arrive again at "a'") it would be irrational to again attempt to solve "a'" using the same steps and methods. It would be irrational because to proceed in this way would produce an infinite cycling around the subproblem "a'". So this second instance of "learning" might be better characterized as an aspect of "rational solution seeking" and not as learning.

The third instance of "learning" appears questionable because L.T. has no means available to determine if such learning is deleterious or not. Consider an example of how this type of "learning" could be harmful for L.T. Suppose L.T. has proven theorems $t_1 \ldots t_{10}$ using method i and axioms r, s, and t. Now suppose L.T. is attempting to prove a new theorem t_{11}. Let us further suppose that it is attempting to prove t_{11} using method i. Because of past experience L.T. will first try axioms r, s, and t in connection with method i. Now it may happen that using method i L.T. is able to prove t_{11} using r, s, and t only after an extensive number of steps, or L.T. may not be able to prove t_{11} using r, s, t at all. In either case a more economical proof may have existed using a different axiom, say g. The reason for this is that $t_1 \ldots t_{10}$ had certain idiosyncratic characteristics whereas t_{11} did not. But because of its "learning" L.T. is forced to use r, s, and t first in attempting to prove t_{11} by method i whether or not such use is economical or effective. Indeed, since L.T. will not attempt to solve a problem using a second way once it has solved a problem in at least some way, this "learning" may force L.T. to attempt proofs of extensive length when other more elegant proofs could be available if L.T. hadn't "learned" in this third way. In short, this third way of "learning" may indeed be harmful learning. Yet L.T. would have no way of "learning" that this was harmful and would blindly follow procedures that may result in slow, if not ineffective, problem-solving.

Let us now consider the second type of verification for L.T. Assume we have constructed a program that purports to provide a theory of problem-solving for a given individual in a given class of situations. This second type of verification proceeds as follows. We present the individual and L.T. with a wff in propositional logic to prove as a theorem. L.T. then prints out the steps it takes in attempting to prove that the wff is a theorem. We then compare this with the steps taken by the individual in attempting to prove that the wff is a theorem to see if the two "match."

With respect to this second type of verification, several questions arise.

Let us presume that we have presented both L.T. and the given individual with a wff to prove as a theorem. Assume that we have a print-out of L.T.'s efforts in attempting a proof of the wff and a record of the individual's attempt to prove the wff a theorem. Now we wish to verify L.T.'s ability to predict the individual's problem-solving behavior by comparing its print-out to the record of the individual. The first problem that we encounter is what constitutes "goodness of fit." In other words, what measure(s) do we have to determine how closely the individual's problem-solving behavior was predicted by L.T.? What would such a measure look like?

How would we justify the use of this measure over some other measure?

Even if we presume that we have a suitable measure of "goodness of fit," another problem is present. A measure of "goodness of fit" will only tell us to what extent, for a given problem, L.T.'s efforts "matched" that of the given individual's efforts. Let us presume for the moment that this measure is expressed as a per cent where for a given problem, 100% "goodness of fit" means that there was a 'perfect match' between L.T.'s predictions and the given individual's performance. The question is: to verify (or falsify) L.T. as a theory of problem-solving for the given individual, what % of "fit" must L.T.'s prediction have with the individual's behavior, and how many problems must be used in the test? Moreover, the question of verification for L.T. (in the second sense of verification) is further complicated by the following. N.S.S.(31) state that as a theory of human problem-solving, L.T. is highly specific: "it describes one organism in a particular class of situations. When either the situation or the organism is changed, the program must be modified." The question then presents itself: if in verifying (in this second sense of the term) L.T.'s theory, we find, in a particular instance, that the "goodness of fit" between L.T.'s prediction and the individual's performance is not what we expected, do we conclude that L.T. has failed (at least in this instance) to predict the individual's problem-solving behavior, or do we conclude that the "particular class of situations" has changed? What does it mean to say that the particular class of situations has changed? How would one determine if this were really the case?

So, it appears that N.S.S.' second type of verification has as many questionable points as the first type of verification.

Perhaps in an effort to avoid these problems, or perhaps unaware of these problems, some individuals have suggested that a theory of problem-solving such as L.T.'s could be verified by the use of Turing's Test. Originally, Turing (32) proposed this test as an operational way to decide questions of artificial intelligence. Here, the test would not be used to test whether L.T. is "intelligent" or can "think" but simply whether L.T. can adequately predict an individual's problem-solving behavior. According to the people who suggest this technique, (33) Turing's Test would be administered as follows: one first suitably records several performances of the human subject and of the computer program in some common code. The code records, one performance to a sheet, are then placed in a container, mixed up, and drawn at random. If an expert cannot reliably tell which performances were produced by the human subject and which by the computer, the program is judged to be adequate as a problem-solving theory for that individual.

Some might think that this test is basically sound although it involves problems such as what constitutes "an expert." However, it would seem that the real objection to this "test" is not that it has certain problems, but that it is totally irrelevant. Remember that as a theory of human problem-solving, we want L.T.'s program to "accurately" predict an individual's problem-solving behavior given a certain problem (or set of problems). All that Turing's Test can tell us is whether or not L.T.'s predictions are "humanoid", not if they "match" the individual's problem-solving behavior. Indeed, for a given problem, given a recording of the individual's behavior and L.T.'s print-out, it may be impossible to distinguish which was produced by the individual and which was produced by the computer and yet L.T.'s prediction may not be "similar" at all to the individual's performance. So Turing's Test (at least the above version) seems to be irrelevant to the verification (in the second sense) that N.S.S. seek.

Perhaps, one final point should be explored in this discussion of verification. The point is that maybe N.S.S.' efforts are misdirected and that they should regard L.T. as an ideal type; that is, L.T.'s program-theory functions in the same way as an ideal theory in the physical science, such as the theory of gases. In other words, the theory is not invalidated by the fact that its predictions possess no precise exemplification in the individual's performance. The theory gives "ideal" predictions that can only be approximated in reality.

It would seem difficult to defend such a position since as Hempel (34) has pointed out, an ideal construct must satisfy the following conditions if it is to be considered similar to an ideal construct in the physical sciences:

1. The introduction of such a construct into a theoretical context requires specification of a set of characteristics and a set of general hypotheses concerning those characteristics.

L.T.'s program would seem to satisfy this requirement since such program possesses a set of symbols (characteristics) and (for hypotheses) a set of information processes that operate on these symbols.

2. An ideal construct must make definite empirical predictions; no use is made of a ceteris paribus clause.

This would seem to provide a problem for L.T. since, as it has been noted, we have the question as to how one decides whether the prediction is inaccurate or the "class of situations" has changed.

3. The "ideal theory" must be derivable from more comprehensive principles which are well confirmed by empirical evidence.

For L.T., one could argue that more comprehensive principles exist. However, even for this more comprehensive set of principles the problem of what constitutes verification remains.(35)

It seems reasonable to conclude that L.T. could not be accurately construed as an ideal type in the area of human problem-solving.

N.S.S. have stated that (36)"...there will be important qualitative similarities...among the programs used by different organisms in a given situation." What do N.S.S. mean by such a statement? Since they provide no further elucidation of this statement, it seems reasonable to assume that they mean at least that certain problem-solving features will be shared by the programs that represent problem-solving theories for a group of individuals, all in the same situation. In other words, each such program will have processes that exhibit certain problem-solving features (or characteristics) that are common to the features exhibited by the processes of the other programs. How could this happen? It seems reasonable to assume that if there is similarity in the processes of all these programs then they (the programs) will exhibit similar problem-solving features. Why would the processes be similar? It seems reasonable to assume that if the problem-solving behavior is somehow similar for these individuals, the processes (that compose the programs) will be similar. In the absence of further elucidation it seems reasonable to infer that N.S.S. believe that qualitative similarities will exist among the programs used by different individuals in a given situation (ultimately) because they will display "similar" (in some sense) problem-solving behavior in a given situation. If this is indeed what N.S.S. are implying then such a statement is open to considerable doubt for the following reasons.

O. K. Moore and S. B. Anderson [37] performed an experiment that was concerned with determining whether or not groups and individuals differ in their modes of attack on complex rational problems.

Six individuals were used and each was presented with the same ten problems to solve in propositional logic. All conceivable experimental controls were taken. In order to make a comparison between individuals and groups, O. K. Moore and S. B. Anderson used six measures to test similarity of problem-solving behavior: (1) number of solutions, (2) mean time for solution, (3) number of steps taken (where a step is the application of a rule of inference to a logic expression), (4) number of errors (where an error is the instruction to use a certain rule of inference which is not applicable to (can't be used on) a certain logic expression), (5) number of cycles (where a cycle is a series of steps from a given logic expression that eventually takes you back to that same expression), (6) number of repetitions.

The results for the six individuals are shown in the table below.[38] The wide range of responses in each column seem to indicate that their problem-solving behavior in a given situation is not very similar. Thus if N.S.S.' statement that "...there will be important qualitative similarities...among the programs used by different organisms in a given situation" is based on the assumption that there will be similarity of problem-solving behavior for these individuals in a given situation, then more proof of this statement is needed.

Over-All Measures of Performance

	I.#1	I.#2	I.#3	I.#4	I.#5	I.#6
Number of Solutions	5	2	2	2	1	1
Mean Time for Solutions Obtained (Minutes)	9	17	4.25	7.5	5	23
Number of Steps Taken	53	144	55	73	146	51
Number of Errors	6	18	24	18	25	8
Number of Cycles	0	25	7	8	25	0
Number of Repetitions	4	15	1	4	12	0

Note: "I" means "individual"

V L.T. In Perspective

In 1952, Karl Deutsch stated, "the social sciences today perhaps are approaching again a 'philosophic crisis' - an age of re-examination of concepts, methods and interests, of search for new symbolic models and/or new strategies in selecting their major targets for attack." Deutsch had in mind the impact of cybernetics (the theory of communication and control) on the social sciences. In what follows, discussion is directed to an investigation of the extent to which the work of Newell, Simon and Shaw did involve ..."re-examination of concepts, methods and interests,....search for new symbolic models and/or new strategies in selecting their major targets for attack."

In particular, we wish to determine the extent to which the work of N.S.S. involved departures from (1) the general methodological approaches of the social sciences at that time, and (2) the work at that time in the psychology of human thought.

Perhaps the most direct way of illustrating the departure of N.S.S. from the general methodological approaches to the social sciences up to that time is to consider the position of their methodology in the long standing naturalistic vs. phenomenological debate. Now the use of the single term, "naturalistic" and the single term, "phenomenological" should not be construed as implying that a clear and uncontested definition of each approach has been, or ever will be, articulated. Indeed, dispute and confusion over definition of these terms arises not only among advocates of one approach and advocates of the other, but among advocates of the same approach and even among neutrals. Thus, we find Leon J. Goldstein[40] defining "the phenomenological approach" as, among other things, one that is primarily oriented toward description and not theory formation, while Thelma Z. Lavine(41) claims that a phenomenological approach (in particular, Verstchen) could solve certain difficulties in the naturalistic position such as the synthesizing of scientific materials into general theory, and the use of symbolic systems in the construction of scientific theories. Perhaps the phrases, "phenomenological approach" and "naturalistic approach" are not simple but cluster concepts. In any case, this does not concern us seriously. What we seek is merely a general characterization of each approach. M. Natanson(42) provides us with one. According to Natanson, the "naturalistic approach" maintains that the phenomena of the social sciences is qualitatively continuous with that (the phenomena) of the physical sciences. Thus, scientific methods generally are held to be not only adequate for the understanding of social phenomena, but indeed, constitute the paradigm for all inquiry in this field. On the other hand, the phenomenological approach maintains that the phenomena of the social sciences are not qualitatively continuous with those of the natural science. Hence, the methods that should be employed must be different from the methods of the physical sciences. In particular, such methods must take into primary account the intentional structure of human consciousness and place major emphasis on the meaning social acts have for the actors who perform them and who live in a reality built out of their subjective interpretation.

We now ask what form these two approaches have taken in the psychology of thought and how Newell, Simon and Shaw's work relates to them.

Perhaps the best example of the phenomenological approach in the psychology of thought is found in the area of clinical psychology where rigorous prediction and verifications (as found in the natural sciences) is almost non-existent. This is not to say that clinical psychology has not provided a wealth of valuable insights, but it certainly suffers from a sparcity of testable theory stated in operational terms. Similar remarks can be directed, although to a lesser degree, to the Gestaltists and the Wurzburg School.

The naturalistic approach has also been evident in psychology. Psychologists have attempted to preserve formal rigor by retreating to simple dichotomous button-pushing choice situations, to the study of reaction times, or to maze experiments with rats. A vast amount of data and experimental technique for this approach is available. In fact, certain formal theories exist - e.g., stochastic learning theory.(44) This approach lacks appeal for several reasons, not the least of which is the frequent willingness to generalize across species with respect to a mechanical model. The basic question is now: how does the approach of N.S.S. relate to these methodologies?

N.S.S. believe that their approach is able to incorporate both these methodologies. Speaking in reference to the phenomenological and the naturalistic approaches, H. A. Simon has remarked:(45)

> Computers now open up a third course of action that requires no compromise. We can continue to deal with complex verbal behavior, but use the computer to simulate it without first encoding it or forcing it into mathematical form.

This statement expresses N.S.S.' <u>belief</u> that their approach incorporates both methodologies. However, the statement lacks definite mention of how both naturalistic and phenomenological elements are present in their approach. Let us see how and to what extent characteristics of both approaches are present in their methodology.

With respect to naturalistic characteristics, it appears possible to note at least three clear examples.

First, recall N.S.S.' definition of "explanation" of human problem-solving.(46) At their level of theorizing, "an explanation of an observed behavior of the organism is provided by a program of primitive information processes that generate this behavior." Notice that their definition of explanation is wholly behavior-oriented. No mention is made of anything subjective or internal. This _prima facie_ behaviorist orientation is a definite naturalistic characteristic.

Second, notice that one of the goals of N.S.S. simulation is to provide their theory with the ability to make rigorous predictions. Referring to L.T.'s program, N.S.S. state: (47)

> ...it (L.T.) makes many precise predictions that can be tested in detail regarding the area of behavior it is designed to handle. For example, the theory contained in this paper predicts exactly how much difficulty an organism with the specified program will encounter in solving each of a series of mathematical problems, which of the problems it will solve, how much time (up to a proportionality constant) will be spent on each, and so on.

Elsewhere, (48)(49) Newell _et al_. state in reference to theories derived from computer simulation:

> These theories are testable in a number of ways: among them by comparing the symbolic behavior of a computer so programmed with the symbolic behavior of a human subject when both are performing the same problem-solving or thinking tasks.

The difficulties in determining what constitutes "verification" have already been noted. However, this quote of Newell clearly indicates a preoccupation with rigorous prediction of human processes that is strongly naturalistic in perspective.

Finally, note that N.S.S.' work is marked by emphasis on operational definition of terms.

With respect to phenomenological characteristics, it would seem to appear that N.S.S.'methodology doesn't exhibit such characteristics at least in any significant way.

Recall the above statement of Natanson that the phenomenological approach maintains that the phenomena of the social sciences are not qualitatively continuous with those of the natural sciences. Thus the methods that should be employed should be different from that of the physical sciences. In particular, such methods must take into primary account the intentional structure of human consciousness and place major emphasis on the meaning social acts have for the actors who perform them and who live in a reality built out of their subjective interpretations.

We have mentioned earlier (page 25) that one of the problems involved in N.S.S.' approach is that of the identification of problem-solving processes. In short, given problem-solving data for a given individual in a given class of situation, we asked how one would proceed in identifying the problem-solving processes involved. Perhaps, here is where the phenomenological elements of N.S.S.' approach become apparent. Why not seek identification (at least in part) of such problem-solving processes by asking the given individual after he has solved a certain problem or problems to somehow provide us either report or account of the processes that he employed? Notice that this is different from the information we gain by having him talk aloud while solving the problem since that information only tells us what steps (and in what order) he is taking in attempting to solve the problem. The information we now seek is what process or procedure did he employ in generating these and only these steps in the given order.

Moreover, he could introspect not only about the process by which he generated his attempted solution steps but also about whether or not he performed certain steps mentally but, for some reason, did not verbally or graphically report them during the problem-solving period.

Now if N.S.S. utilized introspective reports for these two uses, one could certainly argue that this represented phenomenological elements in their methodology. Indeed, Natanson's remark that "In particular, such methods must take into primary account the intentional structure of human consciousness..." seems to imply that certain mental acts, such as the decision to refrain from something, may produce no noticeable physical behavior and may go unobserved if one simply concentrates on behavior. In our case certain problem-solving steps may be taken mentally, but for some reason (say that of subsequent decision on the given individual's part that the step was wrong or stupid) may not have been verbally or graphically noticeable. The same overt act (in our case the same problem-solving step) may be performed by two individuals but for wholly different intentions or reasons (in our case as a result of different problem-solving processes).

So the question of phenomenological elements in N.S.S.' methodology seem to reduce to the question of whether N.S.S. would utilize introspective reports for either of the above two uses. Although N.S.S. have made no actual statement on this issue, it would appear that they would not make use of such reports for the following reasons. For L.T., N.S.S. never actually attempted the type of verification mentioned earlier (page 42) wherein the output of L.T. for a given problem is compared with the verbal or written record made by an individual solving the same problem. Nevertheless, employing the same methodology N.S.S. did attempt such verification for a later computer program. In comparing the computer output with the given individual's performance (called "human protocol") N.S.S. have remarked:(50)

> ...a number of things appear in the computer trace that have no correspondence in the human protocol - most prominently, the references here in the trace to rules 5, 7, and 8. *We cannot tell* whether these omissions indicate an error in the theory, or *whether the subject noticed the rules in question but failed to mention them aloud*. (Italics mine)

This quote seems to imply that there is no overt serious consideration by N.S.S. of the use of introspective reports to determine whether or not a given individual in attempting to solve a problem performed certain mental acts but did not mention them (verbally or graphemically) for some reason or other.

With respect to the use of introspective reports to determine the processes employed in problem-solving, N.S.S., again, have made no actual statement. However, they have remarked:(51)

> We can, in fact, find a number of attempts in the psychological literature to explain behavior in terms of programs -- or the prototypes thereof. ...Quite recently, and apparently independently, we find the same idea applied by Jerome S. Bruner and his associates to the theory of concept formation. Bruner uses the term "strategy" derived from economics and game theory, for what we have called a "program."

It seems reasonable to interpret this quote as an assertion by N.S.S. of a strong methodological similarity between Bruner's approach to concept formation and their approach to problem-solving. If such a strong methodological similarity does exist, then the following remarks of Bruner seem to indicate that N.S.S.' methodology would seem to consider introspective reports irrelevant.(52)

> Let it be said at the outset that a strategy as we are using the term here does not refer to a conscious plan for achieving and utilizing information. The question whether a person is or is not conscious of his strategy, while interesting, is basically irrelevant to our inquiry. Rather a strategy is inferred from the pattern of decisions one observes in a problem-solver...

In sum, it appears that N.S.S.'s approach is not strongly phenomenological. Thus, any claim that their methodology incorporates both phenomenological and naturalistic elements in any significant way is more a wish than it is a fact.

Another point worth mentioning concerning N.S.S.' methodological approach is that they are aware of the fact that for a given phenomenon, there are levels of explanation. H. A. Simon states:(53)

> The goal in simulating complex human behavior is the same as the goal in simulating neural nets: we wish to explain the behavior. But the information processing theories approach that explanation in stages. They first reduce the complex behavior to symbol manipulating processes that have not,

> as yet, been observed directly in the human brain. The hope, of course, is that when we know enough about these processes, it will be possible to explain them at a still more fundamental level by reducing them to systems of neural events.
>
> When this stage is reached, theories in psychology will begin to resemble theories in genetics and in the bio-physical sciences in their hierarchical structure. At the highest (but least fundamental) level will be information processing theories of overt behavior. At the next level will be neurological theories explaining how elementary information processes are implemented in the brain. At a still more fundamental level will be biochemical theories reducing the neurological mechanisms to physical and chemical terms. Information processing theories of thinking neurological theories, and biochemical theories are complementary, not competitive, scientific commodities. We shall need all three kinds, and perhaps others as well, before we shall understand the human mind."

This realization that there exist levels of explanation is not at all new. However, it is methodologically important in this connection since it fosters the realization that a given range of phenomena may be explained at different levels and that the techniques or means of explanation used are not necessarily the same at all levels. Failure to realize this produces not only useless debate but also unfruitful efforts in trying to use the methods of explanation at one level on another level. Moreover, this problem is especially acute because competent writers in the social sciences (being aware of levels of explanation but nevertheless not mentioning them) lead, wilfully or not, an inexperienced reader to believe that there is only one sound methodological technique for investigating all social phenomena. For example, in writing on the notion of ideal types,(54) Carl Hempel states that the theories developed in analytical economics most closely resembles that of the natural sciences. Now Hempel is surely aware that a given range of economic phenomena can be explained at many different levels using many different means of explanation, yet his failure to state this overtly at least suggests to the inexperienced (perhaps, even experienced) reader that one would do well to emulate in other areas and/or levels of explanation the scientific methods as used in analytical economics.

In short, N.S.S.' acknowledgement of levels of explanation with the attending possibility that different levels require different means of explanation, suggests a real desire to avoid the question of ontological priority of one method over another, and instead views alternate methods as complementary rather than competing.

Earlier we attempted to show in general terms of methodology how N.S.S.' approach incorporated both the naturalistic and the phenomenological approach. It would seem worthwhile to be more specific and examine some of the contemporary psychological theories concerning human thinking and note what advantages, if any, are gained by N.S.S.' technique. This will be accomplished by first outlining two prominent contemporary approaches to human thinking and secondly noting advantages or disadvantages of N.S.S.' approach when compared to each of them.

The first approach that is considered is highly dependent on the use of what is known as factor analysis. The technique of factor analysis is involved and complex but the following example should serve to give the reader some notion of its use:(55)

Assume that 24 different tests, which may be referred to as 'Test A', 'Test B',.....'Test X', respectively, have been administered to each of 200 individuals. The question may be raised as to whether what is measured by Test A is the same as what is measured by Test B or by any of the remainder of the 24 tests. Or the question may be raised as to how many different basic variables are required to account for the variations in performance observed in the 24 tests. The first step in answering such questions is to compute the correlation of every test with every other test. The resulting correlation coefficients form a table which is referred to as a correlational matrix. By a factor analysis of the matrix, one can then determine how many independent (uncorrelated) factors are needed to account for the observed intercorrelations among the 24 tests, and also how heavily each of the tests is loaded (weighted) with each of these factors. For example, only five different factors might be necessary to account for the intercorrelations among the 24 tests; thus, Tests A,B,H, and M might all be found to be loaded rather heavily by Factor I, Tests B,D,J,W, and X by Factor II, etc. Ordinarily each test will be loaded with more than one factor; in other words, only rarely will a test be a pure measure of a single factor. If two or more tests have essentially the same loadings on the same factors, then what is measured by one is essentially the same as what is measured by the other(s). The nature of each factor is inferred from the properties which are common to the tests on which it loads heavily; disagreement sometimes occurs as to just how a particular factor is to be interpreted.

Directing their attention to such questions as those concerning the different kinds of operations or processes involved in human thinking, J. P. Guilford *et al.* recently have been the most active in pursuing this approach. According to Guilford,[56] the psychologist's first step in the factor analytic investigation of a particular domain is the formulation of hypotheses:

> In some area of behavior, such as that of visual perception for example, he might hypothesize that seeing visual depth is a function separate and distinct from all other visual-perception functions. According to the hypothesis, individuals should be expected to differ from one another in ability to deal with tasks involving depth perception. The investigator then sets about developing three or more tests, each of which he thinks should indicate such individual differences and each of which is sufficiently different from others in this group of tests to justify believing that they are not just alternate forms of the same test. At this stage, he has no basis for knowing whether all of the tests do indeed indicate individual differences in the same attribute and, if they do, to what extent they succeed. The investigator will think of other perceptual functions that he thinks are distinct from depth perception and from one another and will develop a few tests for each additional hypothesized factor. He will expect the pattern of intercorrelations among all the tests so developed to tell him, through the operations of factor analysis, which of his hypotheses are supported and which are not.

Guilford has presented a model for the representation of "the structure of intellect." The model is as follows:[57]

As can be seen, one dimension of the model divides factors in terms of the kinds of processes or operations involved. Of the five classes represented on this dimension, memory (involving storage of information) and cognition (including discovery, comprehension or understanding) would not ordinarily be included within the concept of thinking. Divergent production involves the generation of new information from perceived and remembered information, as does convergent production. The distinction between them is that in divergent thinking, a variety of answers is produced, whereas convergent thinking leads to one right answer or to a recognized best answer. Evaluation is concerned with operations for determining whether information which is perceived, remembered or produced meets certain criteria.

In the second dimension of the model, factors are divided in terms of the kinds of products that result from thinking or from cognizing or remembering.

In the third dimension, intellectual factors are classified in terms of kinds of contents. Figurial content is concrete material, information perceived directly through the senses (e.g., visual perception of color) whereas semantic and symbolic include abilities involved in dealing with abstract material; symbolic material is composed of letters, digits and other conventional signs, and semantic material is in the form of verbal meanings or ideas. Although no factor yet discovered falls within this category, the behavioral category is included in the model purely on theoretical grounds to represent the general area sometimes called "social intelligence."

The above model includes cells for which no factors have yet been identified and more than one factor may be found in some cells.

The second approach to human thinking that we wish to consider is that of J. Piaget et al. Piaget's work makes use of what he calls the "clinical method." The "clinical" method[58] is characterized by somewhat flexible questions put to the subject in an attempt to probe his understanding of concrete situations with which he is presented. If the subject does not reach the correct answer initially, the experimenter may try to elicit it by the use of additional questions. Such a procedure is supposed to provide a sort of "testing of the limits" by which one can judge whether the unsuccessful subject is truly incapable of formulating the correct answer. Unfortunately, in presenting his results Piaget does not provide quantitative data but simply provides a number of sample protocols for illustrative purpose. In fact, he often does not indicate even the number of subjects or nature of the sample upon which the conclusions are based.

In order to give the reader a general impression of Piaget's thought, the following brief description of the five general stages from birth to adolescence that he believes are involved in the development of thinking is presented:[59]

I. The Sensori-motor Stage (birth to 18 months or 2 years) precedes the appearance of true thinking. During this period, the infant operates on a pre-verbal level; he is unable to produce mentally any linguistic or non-linguistic images or representations of external events, and as a result there is no cognitive activity which is purely internal. This period is one of rapid change. In fact, Piaget distinguishes some six substages within it.

II. The Pre-conceptual Stage (2 to 4 years) involves the beginning of fully internalized, representational activity. At this time, the child develops the capacity to respond to symbols and signs; for Piaget this capacity is fundamental to the development of both thought and language. Symbols are generally prelinguistic and personal, whereas signs are arbitrary, based on social convention, in other words, linguistic forms. The symbols and signs which now become available do not permit true conceptual activity. Rather, the child can employ only what Piaget calls "pre-concepts", something midway between a class concept and the individual members of the class. The child at this stage cannot decide whether successively encountered instances are the same or different members of a class.

III. The Intuitive Stage (4 to 7 years) is a time of transition in that several developments during it come to fruition only in the stage which follows. One clear improvement at this stage, however, is that the child can now grasp the notion of a class with several similar members. But he cannot yet manipulate or coordinate classes in thought dealing with the relationships among them. The child cannot see a situation as a set of causal relations but only as a perceptual configuration, together with the possible modifications which he could produce by his own actions of pushing, picking up, etc. He can carry out simultaneously in thought only those manipulations which could be performed simultaneously in overt action.

IV. The Stage of Concrete Operations (7-8 to 8-11 years) is characterized by the appearance of true operations. Operations are actions which are "internalizable, reversible, and coordinated into systems which are characterized by laws which apply to the system as a whole."

V. The Stage of Formal Operations (11 or 12 to 14 or 15 years) continues the trend, important in earlier stages, of the subordination of reality to possibility, and, for the first time, possibility becomes more important in thinking than reality. Thinking at this stage is characterized by the fact that the individual now deals not only with objects, classes, and simple relations, but with verbal statements about these elements, with propositions and relations between propositions.

Let us now make a comparison of Guilford and Piaget with N.S.S.' work.

As mentioned earlier the factor analytic approach of Guilford is concerned with searching for the smallest number of factors which will account for the intercorrelations among the empirically measured variables. Although this approach provides for well controlled data gathering, clear specification of the operations defining the constructs, and the testing of hypotheses, several limitations present themselves.

First, since a single investigation involves the administration of a fairly large number of tests to each of a large number of subjects, the terms which can be allotted per item is very small. Since this results in tests that include only very short problems, the processes important in solving longer and more complex kinds of problems may escape discovery by factor analysis. The approach of N.S.S. is not limited in this way.

Secondly, the factor analysis approach provides a list of what may be considered to be elementary processes in thinking. However, it provides no account of how these processes fit together in solving a problem or making a decision. We have mentioned earlier that N.S.S.' approach is based on the notion of a complex information processing system, where "complex" meant: a) there is a large number of different kinds of processes, all of which are important, although not necessarily essential, to the performance of the total system; b) the uses of the processes are not fixed and invariable, but are highly contingent upon the outcomes of previous processes and on information received from the environment; c) the same processes are used in many different contexts to accomplish similar functions towards different ends, and this often results in organizations of processes that are hierarchical, iterative and recursive in nature.

Now, at first glance this might tempt one to indicate that N.S.S.' approach also suffers from the same difficulty as Guilford's. However, although the processes are not fixed and invariable, the meta-process that determines which processes are to be used is fixed. This meta-process is simply called the executive routine (as we mentioned earlier) of a computer program. Thus the approach of N.S.S., in contrast to the factor analysis technique, always provides a complete and unambiguous statement of how the elementary processes of the theory fit together.

The above problem seems to be present in a slightly different form in the work of Piaget. We have mentioned and briefly described the five stages that are involved in the development of thinking. However, there is no indication of where or how the processes operating at one stage would give way to the processes of the next stage. In fact, since Piaget uses the concept of "stage" in several ways, in particular in the sense in which successive stages are involved are seen as involving "operation of different rule-systems" it seems reasonable to conclude that a program written at one stage would be essentially different from one written to simulate thinking at another stage.(60)

In addition to this, Piaget's "clinical" method of obtaining data is not well controlled. In contrast, N.S.S.' approach (at least in data gathering) is simple and direct - the data is obtained from a human subject who is asked to use scratch paper as he works a problem, or to think aloud.(61)

Finally, we note that the constructs employed by Piaget are not clearly defined whereas N.S.S.' work is characterized by the clarity of operational definitions.

VI Conclusions

Karl Deutsch has stated[62]:

> The history of many fields of science shows a characteristic pattern. There is a time in which the science goes through a philosophic stage in its development; the emphasis is on theory, on general concepts and on the questioning of the fundamental assumptions and methods by which knowledge has been accumulated.

N.S.S. believe that their approach to human problem-solving will prove fruitful in the development of theories of human learning, perception and concept formation.[63]

Now, if we consider psychology to be a science, then N.S.S.' attempt to develop a theory of human problem-solving based on the use of computer programs would seem to constitute a new philosophic stage for the psychology of human thought. We have shown how their work differs from other attempts to develop an adequate theory of thinking. If one were ever to try to cite the most distinctive feature that distinguishes their approach from any of their predecessors it would seem to be the fact that a theory according to their strategy would maintain scientific rigor without having to limit itself to treating only certain variables.

H. A. Simon has remarked:[64]

> How shall we, for example, characterize the data from a laboratory study of human problem-solving in order to make these data amenable to mathematical and numerical analysis? We can count the number of problems a subject solves in a given time, and assign scores to batteries of problems on the basis of such counts. We can tally numbers of errors of various kinds. But the numbers we obtain in these ways are pale shadows of the subject's actual behavior -- particularly his verbal behavior if he thinks aloud while solving the problem.

It would seem that N.S.S.' approach enables one to treat complex verbal behavior with rigor without first encoding it or forcing it into mathematical form. This fact is, perhaps, the keynote of the new perspective N.S.S.' approach offers to the psychology of thought.

Although the above seems to be true, we have, nevertheless, noted that many problems exist concerning the work of N.S.S. How do these problems relate to this supposed new perspective on the psychology of thought? Consider another quote of Deutsch:[65]

> Philosophic stages in the development of a particular science are concerned with strategy; they attain the targets, or they accumulate experience indicating that the targets cannot be taken in this manner and that the underlying strategy was wrong.

Using Deutsch's terminology, we can answer the above question as follows. We have attempted to provide a thorough exhibition of N.S.S.' approach to the area of human thought (problem-solving in particular) and we have sought to illustrate that their approach did indeed constitute a new philosophic stage in the psychology of human thought. Yet, if their approach is to pass the test of the empirical stage, that is, if the application of this approach is to actually provide us with any new and useful knowledge, the problems we have enumerated must first be resolved. If this is not possible, we must conclude that "the targets cannot be taken in this manner and that the underlying strategy was wrong."

FOOTNOTES

(1) Newell, A., J. C. Shaw and H. A. Simon, 1957, "Empirical Explorations with the Logic Theory Machine," Proceedings of the Western Joint Computer Conference, p. 114.

(2) For other definitions of "heuristic" see:

(a) Gelernter, H., 1959, "Realization of a Geometry Theorem-Proving Machine," Proceedings of an International Conference on Information Processing, p. 135.

(b) Cooper, D. C., 1966, "Theorem-Proving in Computers," Advances in Programming and Non-Numerical Computation, p. 163.

(c) Slagle, J., 1961, "A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus," Computers and Thought, p. 142.

(3) Kleene, S. C., 1967, Mathematical Logic, p. 247.

(4) Newell, A., J. C. Shaw and H. A. Simon, 1962, "The Processes of Creative Thinking," Contemporary Approaches to Creative Thinking, p. 74.

(5) Ibid., p. 75.

(6) Ernst, G. W. and A. Newell, 1966, Generality and G.P.S., Doctoral Dissertation, Carnegie Institute of Technology.

(7) Slagle, J., 1961, A Computer Program for Solving Problems in Freshman Calculus, Doctoral Dissertation, Massachusetts Institute of Technology.

(8) Simon, H. A. and A. Newell, 1956, "Models: Their Uses and Limitations," The State of the Social Sciences, pp. 68-69.

(9) Gardner, Davin A., 1968, The Use of Heuristics in Problem-Solving, Master's Thesis, Lehigh University, p. 50.2.

(10) Newell, A. and H. A. Simon, 1956, "The Logic Theory Machine: A Complex Information Processing System," I.R.E. Transactions On Information Theory, pp. 61-63, 67-74.

It is obvious from the above that this paper was published under the names of Newell and Simon only. However, since this and other papers on L.T. cited throughout this thesis were the result of collaboration among Newell, Simon and Shaw, these papers are treated as if authored by all three.

(11) Newell, A., J. C. Shaw and H. A. Simon, 1957, Op. cit, pp. 218-239

(12) Newell, A., J. C. Shaw and H. A. Simon, 1958, "Elements of a Theory of Human Problem-Solving," Psychological Review, Vol. 65, pp. 151-166.

(13) Newell, A. and H. A. Simon, 1956, Op. cit., p. 62.

(14) Newell, A., J. C. Shaw and H. A. Simon, 1958, Op. cit., pp. 151-152.

(15) Ibid., pp. 158-159.

(16) Hanson, N. R., 1958, Patterns of Discovery, p. 70.

(17) Ibid., p. 13.

(18) Wittgenstein, L., 1958, Philosophical Investigations, G. M. Anscombe (trans.), p. 193.

(19) Hanson, N. R., 1958, Op. cit., p. 15.

(20) Ibid., p. 15.

(21) Ibid., p. 17.

(22) Cannonito, Frank B., 1962, "The Gödel Incompleteness Theorem and Intelligent Machines," Spring Joint Computer Conference, p. 71.

(23) The point of this illustration is that a theory of problem-solving for a given individual must be capable of predicting the individual's performance when any wff is presented as a problem. Clearly, the number of such wffs is denumerably infinite.

(24) The objection raised here is similar to the position maintained concerning naturalism by M. Natanson in "A Study in Philosophy and the Social Sciences," 1963, Philosophy of the Social Sciences, pp. 271-285. Natanson claims that the naturalistic approach to the social sciences cancels out the possibility of self-inspection by its own claim that natural science provides the essential method for stating and evaluating philosophical claims.

(25) Newell, A., J. C. Shaw and H. A. Simon, 1958, Op. cit., pp. 151-152.

(26) Newell, A. and H. A. Simon, 1956, Op. cit., p. 159, p. 161.

(27) Wittgenstein, L., 1958, Op. cit., pp. 94-95.

(28) For a definition and general discussion of Turing Machines see:
Singh, J., 1966, *Great Ideas In Information Theory, Language and Cybernetics*, pp. 184-204.

(29) Aune, B., 1967, *Knowledge, Mind and Nature*, pp. 100-101.

(30) N.S.S. seem to admit to this requirement. See:
Newell, A., 1962, "Learning, Generality and Problem-Solving," *Proceedings of IFIP Congress 62*, p. 408.

(31) Newell, A., J. C. Shaw and H. A. Simon, 1958, *Op. cit.*, p. 151.

(32) Turing, A. M., 1950, "Computing Machinery and Intelligence," *Mind*, Vol. LIX, No. 236.

(33) Employing the same methodology but with respect to a different computer program, N.S.S. seem to advocate this approach. See:
Newell, A. and H. A. Simon, 1961[a], "The Simulation of Human Thought," *Current Trands in Psychological Theory*.

(34) Hempel, C. G., 1963, "Typological Methods in the Social Sciences," *Philosophy of the Social Sciences*, p. 230.

(35) See:
Newell, A., J. T. Shaw and H. A. Simon, 1962, *Op. cit.*

(36) Newell, A., J. C. Shaw and H. A. Simon, 1958, *Op. cit.*, p. 152.

(37) Moore, O. K. and S. B. Anderson, 1955, "Search Behavior In Individual and Group Problem Solving," *American Sociological Review*, Vol. 19, pp. 702-714.

(38) *Ibid.*, p. 708.

(39) Deutsch, K., 1952, "Communication Theory And Social Science," *American Journal of Orthopsychiatry*, Vol. XXII, No. 3, p. 469.

(40) Goldstein, L. J., 1963, "The Phenomenological and the Naturalistic Approaches to the Social," *Philosophy of the Social Sciences*, p. 287.

(41) Lavine, T. Z., 1963, "Note to Naturalists on the Human Spirit," *Philosophy of the Social Sciences*, pp. 259-260.

(42) Natanson, M., 1963, *Op. cit.*, pp. 272-273.

(43) Johnson, D. M., 1955, *The Psychology of Thought and Judgment*.

(44) Bush, R. R. and F. Mostiller, 1955, *Stochastic Models for Learning*.

(45) Simon, H. A., 1961, "Modelling Human Mental Processes," *Western Joint Computer Conference*, p. 112.

(46) Newell, A., J. C. Shaw and H. A. Simon, 1958, Op. cit., p. 151.

(47) Ibid., p. 152.

(48) Newell, A. et al, 1961, Computer Simulation of Human Thinking and Problem Solving, The Rand Corporation, p. 4.

(49) See also the remarks in:
Newell, A., J. C. Shaw and H. A. Simon, 1958, Op. cit., p. 152.

(50) Newell, A. and H. A. Simon, 1961[b], "Computer Simulation of Human Thinking," Science, Vol. 134, p. 2016.

(51) Newell, A., J. C. Shaw and H. A. Simon, 1958, Op. cit., p. 153.

(52) Bruner, J. S., J. Goodrow and G. Austin, 1956, A Study of Thinking, pp. 54-55.

(53) Simon, H. A., 1961, Op. cit., p. 113.

(54) Hempel, C. G., 1963, Op. cit., p. 228.

(55) For a more complete description, see:
Harmon, H. H., 1960, Modern Factor Analysis.

(56) Guilford, J. P., 1960, "Basic Conceptual Problems In the Psychology of Thinking," Fundamentals of Psychology: The Psychology of Thinking, p. 8.

(57) Ibid., p. 10.

(58) Taylor, Donald W., 1963, "Thinking," Theories in Contemporary Psychology, p. 477.

(59) Ibid., pp. 482-484.

(60) For an extended discussion of this, see:
Kessen, W., 1962, "Stage and Structure in the Study of Children," The Thought of the Child.

(61) Newell, A. and H. A. Simon, 1961[b], Op. cit., p. 2012.

(62) Deutsch, K., 1952, Op. cit., p. 469.

(63) Newell, A., J. C. Shaw and H. A. Simon, 1958, Op. cit., p. 166.

(64) Simon, H. A., 1961, Op. cit., p. 112.

(65) Deutsch, K., 1952, Op. cit., p. 469.

BIBLIOGRAPHY

Aune, Bruce. Knowledge, Mind, and Nature. New York: Random House, 1967.

Bruner, J. S., Goodrow, J. and Austin, G. A Study of Thinking. New York: Wiley, 1956.

Bush, R. R. and Mosteller, F. Stochastic Models for Learning. New York: Wiley, 1955.

Cannonito, Frank B. "The Gödel Incompleteness Theorem and Intelligent Machines," Spring Joint Computer Conference. 1962, pp. 71-77.

Cooper, D. C., "Theorem-Proving in Computers," Advances in Programming and Non-Numerical Computation. 1966, pp. 155-182.

Deutsch, Karl. "Communication Theory And Social Science," American Journal of Orthopsychiatry. July 1952, Vol. XXII, No. 3, pp. 469-483.

Ernst, G. W. and Newell, A. Generality and G.P.S., Doctoral Dissertation, Electrical Engineering Department, Carnegie Institute of Technology, 1966.

Gardner, Darien A. The Use of Heuristics in Problem-Solving. Master's Thesis, Information Science Department, Lehigh University, 1968.

Gelernter, H. "Realization of a Geometry Theorem-Proving Machine," Proceedings of an International Conference on Information Processing. 1959, pp. 273-282.

Goldstein, L. J. "The Phenomenological and the Naturalistic Approaches to the Social," Philosophy of the Social Sciences. 1963, pp. 286-301.

Guilford, J. P. "Basic Conceptual Problems In the Psychology of Thinking," Fundamentals of Psychology: The Psychology of Thinking. 1959, pp. 6-21.

Hanson, N. R. Patterns of Discovery. New York: Cambridge University Press, 1965.

Harmon, H. H. Modern Factor Analysis. Chicago: University of Chicago Press, 1960.

Hempel, C. G. "Typological Methods in the Social Sciences," Philosophy of the Social Sciences. 1963, pp. 210-230.

Johnson, D. M. The Psychology of Thought and Judgment. New York: Harper, 1955.

Kessen, W. "Stage and Structure in the Study of Children," The Thought of the Child. Monogr. Soc. Res. Child Develpm. 1962.

Kleene, S. C. Mathematical Logic. New York: Wiley, 1967.

Lavine, Thelma Z. "Note to Naturalists on the Human Spirit," Philosophy of the Social Sciences. 1963, pp. 250-261.

Moore, O. K. and Anderson, S. B. "Search Behavior In Individual and Group Problem-Solving," American Sociological Review. 1955, Vol. 19, pp. 702-714.

Natanson, M. "A Study in Philosophy and the Social Sciences," Philosophy of the Social Sciences. 1963, pp. 271-285.

Newell, A. "Learning, Generality and Problem-Solving," Proceedings of IFIP Congress 62. 1962, pp. 407-412.

Newell, A. et al. Computer Simulation of Human Thinking and Problem Solving. Santa Monica: The Rand Corporation, 1961.

Newell, A., Shaw, J. C. and Simon, H. A. "Empirical Explorations with the Logic Theory Machine," Proceedings of the Western Joint Computer Conference. 1957, pp. 218-239.

Newell, A., Shaw, J. C. and Simon, H. A. "Elements of a Theory of Human Problem-Solving," Psychological Review, Vol. 65, pp. 151-166.

Newell, A., Shaw, J. C. and Simon, H. A. "The Processes of Creative Thinking," Contemporary Approaches to Creative Thinking. 1962, pp. 63-117.

Newell, A. and Simon, H. A. "The Logic Theory Machine: A Complex Information Processing System," I.R.E. Transactions On Information Theory. 1956, pp. 61-79.

Newell, A. and Simon, H. A. "The Simulation of Human Thought," Current Trends in Psychological Theory. 1961, pp. 480-491.

Newell, A. and Simon, H. A. "Computer Simulation of Human Thinking," Science. 1961, Vol. 134, pp. 2011-2017.

Simon, H. A. "Modelling Human Mental Processes," Western Joint Computer Conference. 1961, pp. 66-83.

Simon, H. A. and Newell, A. "Models: Their Uses and Limitations," The State of the Social Sciences. 1956, pp. 66-83.

Singh, J. Great Ideas In Information Theory. New York: Dover, 1966.

Slagle, J. A Computer Program for Solving Problems in Freshman Calculus. Doctoral Dissertation, Massachusetts Institute of Technology, 1961.

Slagle, J. "A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus," Computers and Thought. 1961, pp. 180-195.

Taylor, Donald W. "Thinking," Theories in Contemporary Psychology. 1963, pp. 475-493.

Turing, A. M. "Computing Machinery and Intelligence," Mind. 1950, Vol. LIX, No. 236.

Wittgenstein, L. Philosophical Investigations. G. M. Anscombe (trans.) New York: Macmillan, 1958.

Vita

John Patrick Coyne, son of John Thomas and Mary Coyne, was born in New York, New York, on December 30, 1945. He was graduated from Power Memorial Academy in New York, New York, in June 1963 and received a Bachelor of Science Degree from Iona College in June 1967.