DOCUMENT RESUME

ED 113 196 SE 019 757

AUTHOR Scandura, Joseph M.; And Others
TITLE Higher Order Rule Characterization of Heuristics for Compass and Straight Edge Constructions in Geometry.
PUB DATE Apr 75
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975); Best copy available; Occasional marginal legibility

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Algorithms; Artificial Intelligence; *Cognitive Processes; *Geometry; Information Processing; *Mathematics Education; *Problem Solving; *Research; Secondary Education; Thought Processes
IDENTIFIERS *Geometric Constructions; Polya (George)

ABSTRACT

The research reported in this paper was designed to analyze the incidence of use of higher-order rules by students solving geometric construction problems. A carefully selected set of construction problems was subjected to rigorous a priori analysis by mathematics educators to determine what basic and second-order rules might be used by able high school students in their solution. Categories of problems analyzed include: patterns of two loci, patterns of similar figures, combined two loci and similar figures, patterns of auxiliary figures, and patterns of loci, similar figures, and auxiliary figures. The analysis was successful in making more precise the heuristic approach of George Polya. Overall, the viability of this method of analysis was demonstrated. The authors cite some limitations of the study and future directions for their work. (SD)

HIGHER ORDER RULE CHARACTERIZATION OF HEURISTICS FOR COMPASS AND STRAIGHT EDGE CONSTRUCTIONS IN GEOMETRY[1]

Joseph M. Scandura, John H. Durnin,[2] and Wallace H. Wulfeck II
University of Pennsylvania

According to Polya (1962), perhaps the greatest value to be gained from the study of mathematics is the ability to solve problems. In spite of its importance, however, relatively little is known about how to teach people to solve problems, or how to program computers to do so. Specifically, one of the great mysteries of our time is why some problem solvers (human or computer) succeed on problems for which they have all of the necessary component skills (operators) whereas others fail.

In dealing with this question most research in AI has been concerned with the construction of powerful computer programs which can solve more or less diverse classes of complex problems. In computer simulation an attempt is made to also parallel human performance on such problems. In general, such systems (e.g., Newell & Simon, 1972; Minsky & Papert, 1972) have been comprehensive in scope; they have been concerned with problem definition (the construction of subgoals), memory, the derivation of solution procedures, and the use of such procedures.

The present research has adopted a somewhat different strategy. It seeks understanding by dealing separately with the various aspects of problem solving (e.g., the derivation of solution procedures). In particular, this research is concerned with the specification and testing of general, potentially useful heuristics for constructing procedures for solving compass and straightedge construction problems in geometry. The research also was concerned with developing and determining the feasibility of a general method by which heuristics may be identified in arbitrary problem domains.

One general point of departure was Polya's (1962) work on heuristics for geometry construction problems. These heuristics are purposely cast in a form designed to parallel human thought processes in much the same way as are such general heuristics as means-ends analysis (e.g., Newell & Simon, 1972). Human processing presumably is highly efficient in many situations, and the importance of paralleling human processing in AI, as well in computer simulation, has become increasingly well recognized as a means of significantly reducing processing time. Winston (1972), for example, has noted how constraining syntactic procedures to reflect underlying semantics in the recognition of block scenarios can drastically reduce the number of possibilities that must be considered.

In spite of the broad acclaim for Polya's work generally, however, and the intrinsic support for his notion of heuristics specifically, it sometimes has been difficult to capitalize on these ideas as fully as might be desired. Although often useful, his heuristics frequently are little more than general hints, and leave much to be desired insofar as pinpointing what a human or computer must know in order to solve specific kinds of problems. In order to lend themselves to technological treatment, heuristics must be transformed or incorporated into strictly mechanical procedures that can be more or less readily implemented on computers. Ideally, one might desire reduction of heuristics to algorithms; witness the alpha-beta "heuristic" (e.g., Nilsson, 1971).

Since heuristics tend to be (problem) domain specific, the potential value of more or less general and systematic methods for specifying heuristics in arbitrary problem domains seems fairly clear. Our approach to this problem was designed to be compatible with Scandura's (1973) theory of structural learning, and is an extension of a method used earlier by Ehrenpreis and Scandura (1972). That portion of the theory with which this research is most concerned has been shown empirically to reflect the behavior of individual subjects in particular situations where problem definition and

1. This research was supported by National Science Foundation Grant GW 6796 to the first author. An unabridged version of this paper is available from Joseph M. Scandura, 3700 Walnut Street, University of Pennsylvania, Philadelphia, Pa. 19174.

2. Now at Villanova University.

with some empirical verification (Scandura, 1973a; Scandura, 1973b) to be extendable to situations involving memory and, apparently, also problem definition (Scandura, 1973a, p. 348) and perception (Ch. 5), without essential change. The structure of the theory must be enriched in these cases but without affecting its basic character (i.e., underlying behavior mechanism). This research is based directly on one part of the idealized theory, in particular that part which is concerned with competence--the specification of rule sets which account for classes of problems. In this theory, a rule set is said to account for a class of problems, roughly speaking, if for each problem in the class (1) there is a solution rule (operator) in the rule set which has the problem in its domain and whose range contains the solution to the problem or (2) there is a higher order rule in the rule set which applies to rules in the set and generates a solution rule. In such a rule set, higher order rules correspond to heuristics. (For a more general and formal formulation, which allows for any number of levels of derivation and in which the rules are not in a fixed hierarchy, see Scandura, 1973a, Ch. 5; 1973b.)

It seems unlikely, of course, that algorithmic methods can be found for devising nontrivial rule sets or heuristics. Indeed, as Chomsky (1968) has argued in the case of linguistics, no such method exists for dealing with observables as complex as language. Work in automatic programming, on the other hand, while it is quite far at present from a satisfactory solution, is proceeding as the authors[3] understand it, on the assumption that significant progress in this direction can be made.

In the present research, the task of specifying heuristics is made simpler in at least two ways. First, and most important, the type of competence theory proposed imposes important constraints on the nature of allowable rule sets, and in turn on the form of the heuristics (higher order rules). In particular, higher order rules are assumed to operate on component (lower order) rules to generate integrated problem solution rules (procedures). These rules may simply compose component rules but may also modify them, for example, by generalization or restriction rules (Scandura, 1973a).

Second, restricting the level of analysis to that of flow diagrams, rather than computer programs, makes it natural to represent the constituent operations and decision making capabilities at whatever level seems to most adequately reflect human knowledge rather than at a level predetermined by some programming language. (We do not mean to minimize the importance of devising working programs. In fact, parts of this analysis have been implemented by one of the authors.) While no general assurance can be given with regard to any particular method, it would seem that a method which results in heuristics (and simple operators) that appear consistent with human thought would have a reasonable chance of having general value.

METHOD OF ANALYSIS

Our method of analysis went something as follows. First, we attempted to set some reasonably explicit bounds on the class of geometry construction problems to be considered. In particular, we considered only those problems in or like those of Chapter 1 of Polya (1962).

Our next step was to classify these problems on heuristic-intuitive grounds. Our aim was to place similar problems in the same categories, in accordance with the general form of their solutions. We were one step up in this regard, since Polya had already done part of the categorization for us. All of his problems can be solved according to some variant or combination of the three general heuristics he describes: (1) the pattern of two loci, (2) the pattern of similar figures, and (3) the pattern of auxiliary figures.

After the various tasks had been classified, we made sure that the domains and ranges of each task were fairly explicit. Then we identified explicit procedures for solving each type of task. Care was taken to insure that these procedures reflected our

3. The authors do not profess to be experts in AI or in computer simulation as such, but rather in the adjacent and we think complementary domain of structural psychology which is, in our view, considerably broader than contemporary cognitive and information processing psychology.

intuitions as to how intelligent high school students might go about solving the problems. In some cases it was possible at this point to subclassify some of the tasks.

The most critical step was to identify general parallels among the procedures developed for the sampled problems within each of the various classifications, and even more important to devise higher order rules (operator combination methods) which realized these parallels as relatively formal, but still general, procedures. The higher order rules so identified (together with the component lower order rules on which they act) provided a general basis for constructing solution rules for the sampled problems.

Then we attempted to refine the resulting higher order rules with regard to specific sampled problems. This was done systematically; where a higher order rule failed to yield an adequate solution rule for a sampled problem, appropriate modifications in the higher order rule were made. A serious attempt also was made to insure that the higher order rules were compatible with human knowledge.[4]

PATTERN OF TWO LOCI

Our first step was to select a broad sampling of two-loci problems and to devise procedures for solving each. For example, consider the problem: "Given a line and a point not on the line, and a radius R, construct a circle of radius R which is tangent to the given line and which passes through the given point." This problem can be solved according to the following procedure: "Construct the locus of points at distance R from the given point; construct the locus of points at distance R from the given line; construct a circle using the intersection point of the two loci as center, and the distance R as radius."

This solution rule clearly involves the pattern of two loci. In this case, as with all of the problems in Polya's first category, the tasks may be characterized according to the form of their solution procedures: two loci are determined one after the other; the point of intersection of these loci in turn makes it possible to construct the goal figure.

Further analysis of the class of two-loci problems, however, revealed differences in the ways problems are solved. In many solution rules, for example, like the example above, the two loci can be found independently, in either order. Furthermore, at no point in the course of applying the solution rule is it necessary to measure a distance. Some form of distance measurement, however, is required with other tasks. Some of the sampled tasks require measurement in order to construct the goal figure; the solution rule for another problem involves measurement before the second locus can be found. In still another task, one of the loci is actually given, or equivalently, can be thought of as obtained by applying an identity rule. The goal figure in still another task is simply the point of intersection of the two loci.

An initial characterization

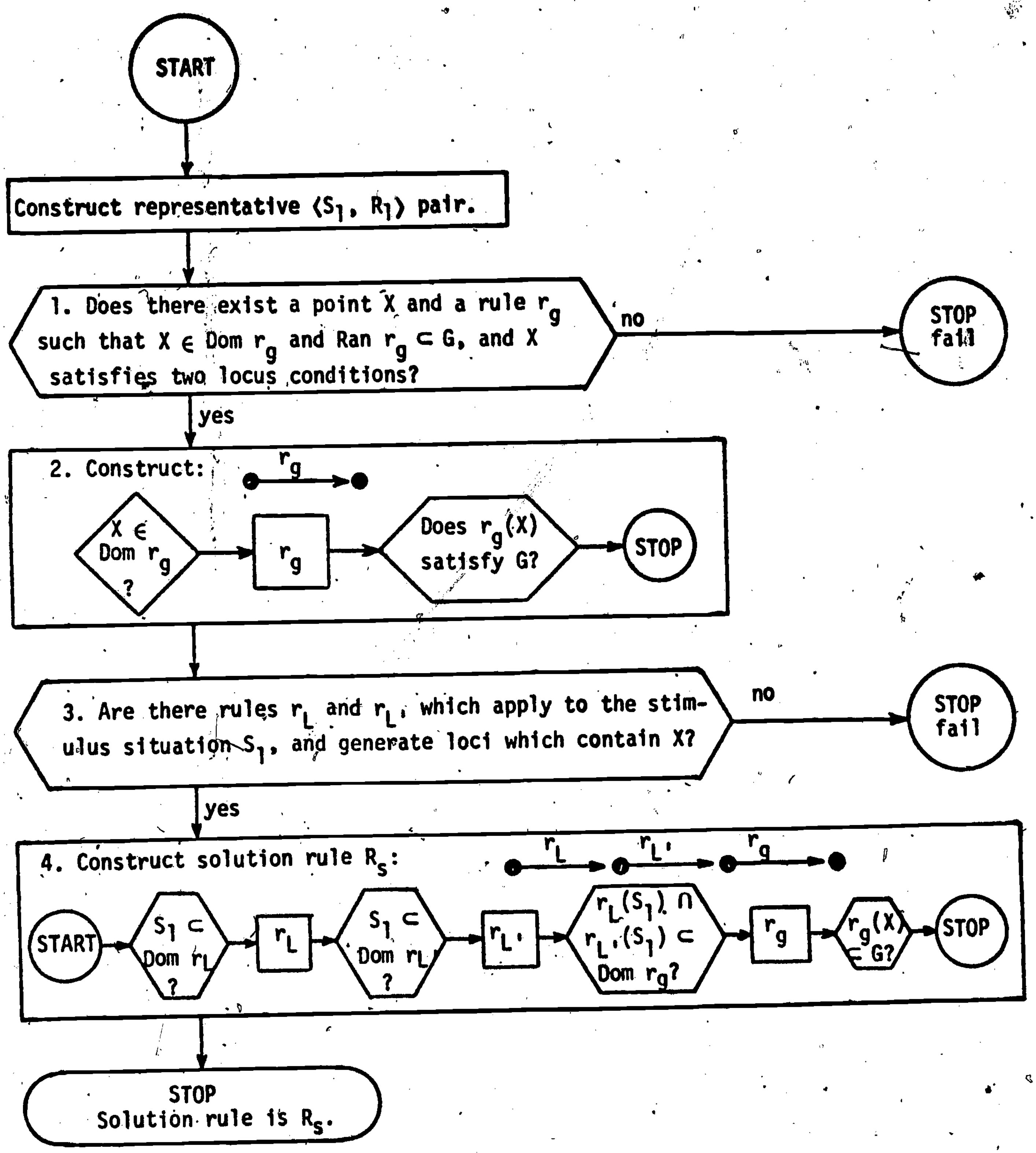
As a first step in characterizing a two loci higher order rule, we systematically went through the various solution rules for the pattern of two-loci tasks and identified all of the different component rules that appeared in our sample problems either (1) in constructing one of the loci, or (2) in constructing a goal figure. The lower order rules we identified were mostly common constructions (e.g., perpendicular bisector, circle, parallel line). Some of the lower order component rules were used to construct a needed locus, others were involved in constructing goal figures, and some served both functions.[5]

The higher order rule in Figure 1 shows schematically how the various solution rules may be constructed from the component rules.

4. See Appendix A.

5. Lists of the component rules involved in our analyses are available in the unabridged report. See footnote 1.

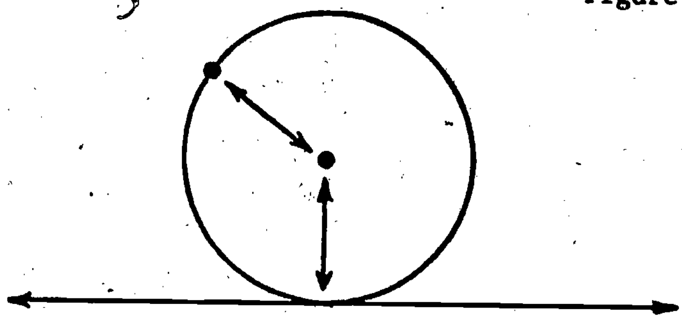
Figure 1


START
Construct representative $\langle S_1, R_1 \rangle$ pair.
1. Does there exist a point X and a rule r_g such that $X \in$ Dom r_g and Ran $r_g \subset G$, and X satisfies two locus conditions?
no
STOP fail
yes
2. Construct:
r_g
$X \in$ Dom r_g ?
r_g
Does $r_g(X)$ satisfy G?
STOP
3. Are there rules r_L and $r_{L'}$ which apply to the stimulus situation S_1, and generate loci which contain X?
no
STOP fail
yes
4. Construct solution rule R_s:
r_L
$r_{L'}$
r_g
START
$S_1 \subset$ Dom r_L ?
r_L
$S_1 \subset$ Dom $r_{L'}$?
$r_{L'}$
$r_L(S_1) \cap r_{L'}(S_1) \subset$ Dom r_g?
r_g
$r_g(X) \subset G$?
STOP
STOP
Solution rule is R_s.

The higher order rule in Figure 1 applies to the problem (i.e., the stimulus situation, S_o) and to the goal (G) itself, as well as to the lower order component rules.[6]

First, an arbitrary representation $\langle S_1, R_1 \rangle$ analogous to the solved problem is constructed. In our illustrative task, a sketch like Figure 2 would serve this purpose.

Figure 2



Note that constructing such a representation is not the same either as solving the problem, or as constructing a solution rule for the problem. The sketch in Figure 2, for example, can easily be generated by first drawing an arbitrary circle, then drawing an arbitrary line tangent to it, and placing an arbitrary point on it. More generally, an arbitrary representation (R_1) of the goal figure (R_o) is constructed first. Only then is a representation (S_1) of the information given in the stimulus situation (S_o) constructed in relation to the representation of the goal figure. In effect, the first operation on the higher order rule amounts to representing geometrically the meanings of goal situations (i.e., goals plus stimulus situations) by a "sketch," or some equivalent representation.[7]

The second step is the question: "Is there a point X in $\langle S_1, R_1 \rangle$ which satisfies two locus conditions - and, if so, is there a goal constructing rule (r_g) such that point X is contained in the domain of r_g (Dom r_g) and such that the range of r_g (Ran r_g) is contained in the goal, G?"

As shown in Scandura (1973a), decision making capabilities can be characterized as partitions on a class of input situations; in the present case, each representation $\langle S_1, R_1 \rangle$ either contains a point X which satisfies two locus conditions or it does not. If it does satisfy two such conditions, then the next operation involves forming the rule consisting of (1) a decision which asks whether there is a point X in the domain of r_g which satisfies two locus conditions, (2) the rule r_g, and (3) stop.

Next, the available component rules are tested to see whether there are two of them which apply to the represented stimulus (S_1) and generate loci which contain the point X. Given that such locus rules exist, the next operation constructs the solution rule R_s in which first one locus rule r_L is applied (after testing to see whether the stimulus situation is in its domain), when the other $r_{L'}$, and finally the goal construction rule r_g.[8]

A more rigorous analysis

This level of description is sufficient to give one an intuitive feeling for how the higher order rule operates. But the rule is ambiguous, especially for computer implementation purposes. In the first decision making capability, for example, it is not clear just what constitutes a locus condition. Similarly, in the second decision making

6. See Appendix B.
7. See Appendix C.
8. See Appendix D.

capability the notion of a rule applying to a stimulus situation is something less than precise.

Closer perusal of the individual tasks made it possible to overcome these ambiguities. In many cases, the desired point X is a given distance from one or two given points and/or lines. In the example above, for instance, the point X is a distance R from the given point and from the given line. This suggested the following more rigorous characterization of the first decision making capability:

(1) Does there exist a point X in $\langle S_1, R_1 \rangle$ and a rule r_g such that (X, E) is contained in the domain of r_g where E is a given distance, and the range of r_g is contained in the goal (Ran $r_g \subset G$) such that X is a given distance from one or two given points and/or lines?

A similar analysis suggested reformulating the second decision making capability as:

(2) Is there a rule r_L such that a pair consisting of given points, lines, and/or distances in S_1 is in the domain of r_L (Dom r_L) and such that X is a member of L (i.e., a point on L) where L is contained in the range of r_L ($X \in L \in$ Ran r_L)?

A similar characterization is required for $r_{L'}$.

A higher order rule incorporating these refinements can be used to generate solution rules for many two-loci problems. For example, in the illustrative problem there is certainly a point X in the representation $\langle S_1, R_1 \rangle$ which is at the given distance R from a given point and from a given line in S_1. It is also true that there is an r_g rule which applies to the pair consisting of the point X and the given distance, and whose range consists of circles and is thereby contained in the goal.

Unfortunately, as it stands, the modified higher order rule does not provide an adequate means for characterizing solution rules for other sampled two-loci tasks. In certain tasks, for example, no distance is given. The important requirement in such cases is often that the point X be <u>equidistant</u> from a given pair of elements, points and/or lines, in two different instances (i.e., for two given pairs of elements). Thus, in the tasks, "Inscribe a circle in a given triangle," the desired point X is equidistant simultaneously from two different pairs of sides of the triangle, or equivalently, the point X is equidistant from the three sides.

Still other tasks involve the (lower order) rule for constructing the locus of vertices of an angle of given measure subtending a given line segment. The task, "Given side a of a triangle, the median M_a, and the measure of angle A opposite side a, construct the triangle," is of this type. The locus of vertices, in this case, is an arc but the points on it are not at a fixed distance from any point on the given segment. Nor are the points of the locus equidistant from any two particular points on the line segment.

In order to take these possibilities into account, the decision making capability was generalized so that the point X could be equidistant from a pair of points or lines, or could serve as a vertex of an angle of given measure whose sides subtend (i.e., pass through the end points of) a given segment. Decision making capability (3) was also enriched so that pairs consisting of angle measures and/or segments could be in the domain of a locus rule.

Further, in the problem, "Given three intersecting lines, not all intersecting at a common point, construct a circle which is tangent to two of the lines and whose center is on the third," we have a situation where one of the loci, the line containing the point X, is already given. To handle this possibility we simply assume an "identity" lower order rule, one which identifies a given line as a required locus.

With these modifications, the higher order rule handled almost all of the pattern of two loci tasks we had sampled. We ran into difficulty, however, with another task: "Given two parallel lines and a point between them, construct a circle which is tangent to the two lines and passes through the point." This difficulty involved the second decision making capability (3). There is a pair of lines in the domain of one of the locus rules - one which constructs the locus of points equidistant from the two given parallel lines. The second locus rule, however, requires that we first measure a distance between two parallel lines, one of which is not present in the stimulus S_o until after the first locus rule is applied. That is, we need to determine the distance between one of the parallel lines and the locus of points equidistant from the two given parallel lines. This distance serves as the desired radius.

Application of the higher order rule in this case results in failure at decision making capability (3). Fortunately, it is easy to modify the higher order rule to take this possibility into account. Furthermore, as we shall see, this modification serves an important purpose in dealing with the larger class of construction problems solvable either by the pattern of two loci or by the pattern of similar figures.

Instead of stopping when the second decision fails, we simply add another group of tests (A-C). (A) and (B) duplicate (1) and (2) except that X must satisfy only one specific condition. (C) asks: "Is there one component rule such that a pair of given points and/or lines is in the domain of that rule and is there a locus L such that the point X is part of L and L is contained in the range of r_L?" If the answer to this is no, we stop, but if the answer is yes, we can ask whether there is another locus rule r_L such that the represented stimulus situation S_1, together with the preceding locus $r_L(S_1)$, contains a pair of given points and/or lines that are in the domain or r_L.

A revised higher order rule which incorporates all of these modifications is shown in Figure 3, found on the following page.

In checking this higher order rule we found it to provide an adequate account not only of all of the pattern of two loci problems sampled, but others as well. For example, consider Task A: "Given sides a, b, and c of a triangle, construct the triangle." In this case, application of the higher order rule generates the solution rule. This solution rule involves: (1) application of the rule, "Construct the locus of points at a given distance from a given point," to the end point of one line segment using another side as distance, followed by (2) another application of the rule to the other end point using the remaining side as radius. Then, the triangle rule, "from a point not on a given segment, draw segments to the end-points of the given segment," is applied to the intersection of these two loci to obtain the desired goal figure.

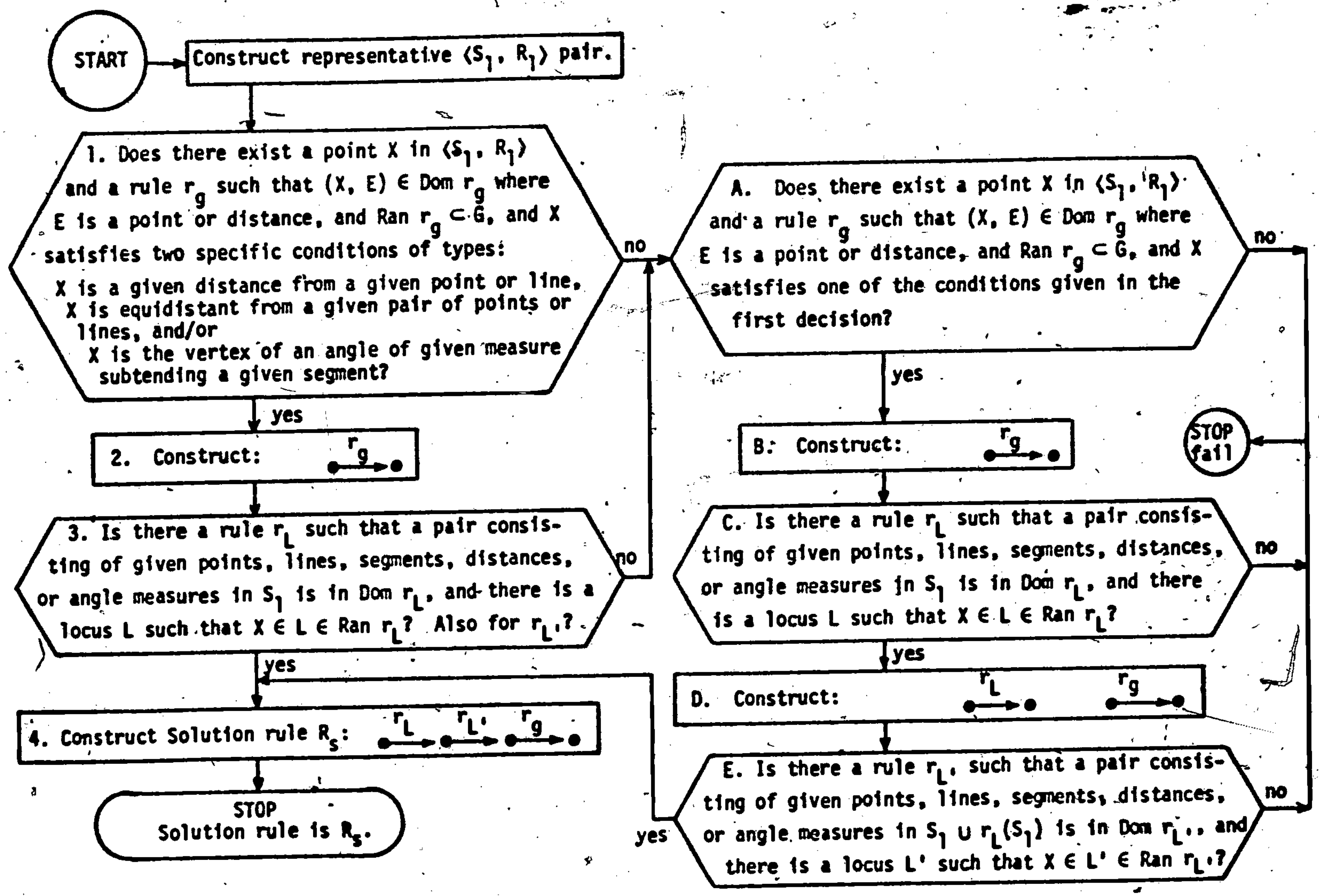
In some cases, of course, different lower order (component) rules were involved. For example, consider task B, "Given two intersecting lines and a point of tangency on one of the lines, construct a circle which is tangent to the two lines and which passes through the given point of tangency." In this case, the locus rule for constructing perpendiculars to lines through points on the given lines had not been required with any of the sampled problems.

Discussion

Aside from the possibility that new two-loci problems may require additional lower order rules, the higher order rule appears adequate. In particular, the higher order rule not only generates solution rules for each of the sampled two-loci problems, but also seems compatible with human knowledge.

As the form of the higher order rule suggests, the component decision making capabilities play a crucial role in deriving solution procedures. These decision making capabilities are designed to reflect the underlying semantics of the problem situations by referring directly to figural representations of semantic information implicit in the problem descriptions. In general, parts of a figural representation $\langle S_1, R_1 \rangle$ will represent the meaning of a task statement and reflect the relation between the given stimulus (S_o) and the goal figure (R_o). Notice that while the relation between S_1 and R_1 will be the same as between S_o and R_o, S_1 and R_1 will not in general be the same as S_o and R_o, respectively.

Figure 3



START
Construct representative ⟨S1, R1⟩ pair.
1. Does there exist a point X in ⟨S1, R1⟩ and a rule rg such that (X, E) ∈ Dom rg where E is a point or distance, and Ran rg ⊂ G, and X satisfies two specific conditions of types:
X is a given distance from a given point or line,
X is equidistant from a given pair of points or lines, and/or
X is the vertex of an angle of given measure subtending a given segment?
yes
no
2. Construct: rg
3. Is there a rule rL such that a pair consisting of given points, lines, segments, distances, or angle measures in S1 is in Dom rL, and there is a locus L such that X ∈ L ∈ Ran rL? Also for rL'?
yes
no
4. Construct Solution rule Rs: rL rL' rg
STOP
Solution rule is Rs.
A. Does there exist a point X in ⟨S1, R1⟩ and a rule rg such that (X, E) ∈ Dom rg where E is a point or distance, and Ran rg ⊂ G, and X satisfies one of the conditions given in the first decision?
yes
no
B: Construct: rg
STOP
fail
C. Is there a rule rL such that a pair consisting of given points, lines, segments, distances, or angle measures in S1 is in Dom rL, and there is a locus L such that X ∈ L ∈ Ran rL?
yes
no
D. Construct: rL rg
E. Is there a rule rL', such that a pair consisting of given points, lines, segments, distances, or angle measures in S1 ∪ rL(S1) is in Dom rL', and there is a locus L' such that X ∈ L' ∈ Ran rL'?
yes
no

For purposes of our analysis, the decision making capabilities were viewed as atomic although they can also be analyzed into more basic components. The first decision making capability in the second two loci higher order rule, for example, involves both a conjunction and disjunction of a number of simpler conditions. This decision making capability could be subdivided, for instance, into the following two decisions: (A) Is there a point X that is a given distance from a given point and/or line? (B) is there a point X equidistant from a pair of given points or lines? Instead of having one decision making capability involving conditions A and B, then, we could have one decision making capability involving A, and a subsequent one, B.[9, 10]

In addition to its purported compatibility with human knowledge, the higher order rule is also sufficiently precise to be mechanizable. One of the authors (Wulfeck) has recently written a program in SNOBOL 4 which uses an intermediate version of the two loci higher order rule (see the unabridged report) to generate solution procedures for many of the problems we sampled. A naming system replaced the figural representation described above (see footnote 7). Routines corresponding to many of the lower order rules (see the appendices of the unabridged report for lists of the component rules) were also written.

Granting the adequacy of the higher order rule for purposes of our analysis, we wish to comment briefly on some limitations in regard to the compatibility of the lower order rules with human knowledge, though the specification of component rules is not our central concern. These limitations are all variants on a common theme: The lower order rules we have identified can be constructed from more basic components. This fact is reflected in at least three ways.

First, many of the simple rules have components in common. Several rules, for example, all involve constructing a locus of points (circle) at some distance from some point. The differences lie in whether or not the distance and/or center points are given directly or must be determined first. The construction rules needed to determine these distances and/or center points are quite basic and are apt to be useful in a wide variety of construction situations. Any reasonable account, designed to deal with a wider variety of problem situations, would undoubtedly include these construction rules directly in the rule set.

Second, certain of the identified lower order rules, particularly the rule for constructing the locus of vertices of an angle of a given measure subtending a given line segment, are complex in themselves and cannot automatically be assumed to be available to many problem solvers.

A third limitation is closely related to the first and was mentioned earlier: the lower order rules are to some degree specific to the tasks we have identified. To some extent this may be unavoidable because there are always certain problems which require "trick" solutions. It would be desirable, of course, to keep this to a minimum. In this regard, it should be emphasized that the simpler the lower order rules the greater the problem solving flexibility.

One way to modify our characterization to handle these limitations would be to "reduce" the lower order rules into their components and, correspondingly, to "enrich" the higher order rule by adding sub-routines for constructing the needed locus, r_L, and goal, r_g, rules.[11] Such rules would correspond to the type of knowledge that a person just having been taught the basic construction rules would need to have in order to generate solution rules directly.

For example, consider the rule: "Determine the distance between a given point and a given line and then construct the locus of points at the obtained distance from the given point." This rule can be divided into two subrules: (1) "Determine the distance between a point and a line," and (2) "Construct the locus of points at a given distance from a given point." To compensate for the reduction in the latter case, the higher order rule could be "enriched" so that more complex r_L and r_g rules can be

9. For a discussion of how new decision making capabilities are learned from simpler ones, see Scandura (1973a).

10. Such refinement may be useful in the assessment of behavior potential (durnin & Scandura, 1973), specifically in increasing the precision of diagnostic testing.

11. See Appendix E.

generated where needed. Specifically, instead of ... when it meets certain prescribed conditions, as we have done so far, we include in the higher order rule a simple sub-routine for combining component lower order rules. Such a sub-routine for example, might select (sub)rules until one is found whose domain includes a pair consisting of a point and a line (e.g., the distance measuring rule (1), and another (e.g., the circle rule (2)) such that its range consists of circles (loci). To make the search more efficient, it is natural to add the requirement that the range of the former be contained in the domain of the latter. After the component rules have been identified, the sub-routine would form the composite of these rules, and finally, would test the composite against the condition in the initial higher order rule.

As attractive as this possibility might appear at first, a little thought suggests its implausibility as a way of modeling human knowledge. This can be seen by noting that all geometric constructions with straightedge and compass are generated by just three basic operations: (a) using a straightedge (e.g., to draw a line, ray, or segment through two given points, or through one point, or intersecting a line, etc.) (b) drawing an arc given a compass set at some fixed radius, and (c) given two points, setting a compass to the distance between those points.

As we have seen, many of the lower order rules are really quite complex. Requiring a higher order rule, designed to reflect human knowledge, to generate such rules from elemental components is unrealistic. It is unlikely that a subject who is only abl to perform the three indicated operations above would also have at his command a rather complex and sophisticated higher order rule. The acquisition of such complex capabilities by naive subjects, whether of a higher or lower order, would almost certainly have to come about gradually through learning, presumably by interacting with problems in the environment.[12]

PATTERN OF SIMILAR FIGURES

Three classes of similar figures problems

The pattern of similar figures problems were analyzed in similar fashion. Again, we began with a broad sampling of problems from Polya (1962). One of the problems identified was, "Given a triangle, inscribe a square in it such that one side of the square is contained in one side of the triangle and the two other opposite vertices of the square lie on the other two sides of the triangle." The second step was to identify a solution rule for each of the problems. For the problem above the solution rule was, "Construct a square of arbitrary size such that one side is contained in the side of the triangle which is to contain the side of the goal square, and such that one vertex is on another side of the triangle. Draw a line through the point of intersection of those two sides of the triangle and through the fourth vertex of the arbitrary square. From the intersection of this line and the third side of the triangle (which is the fourth vertex of the goal square) construct a segment perpendicular to the side of the triangle which is to contain a side of the goal square. Complete the goal square using the length of the perpendicular segment as the length of the sides."

Similar figures problems, like the example task above, may be characterized as those whose solution procedures involve a similarity mapping process: from some center (point) of similarity a figure or set of points is mapped onto another. Further, the solution procedures always involve constructions according to geometric invariants unde similarity mappings, either parallel lines, since parallelism is preserved, or equivalently, "copying" angles, since similarity maps are conformal.

Further analysis of the similar figures problems revealed three relatively distinct classes of solution rules. In the sample problems above, and in other problems i the same class, the solution rules all involve first constructing a square of arbitrary size which is in the same orientation as the desired goal square, and which meets as many of the task conditions as possible. (Rules of this type for constructing similar

---

12. See the section on "future directions".

figures are denoted by r_{gs}.) The second step in each solution rule uses two pairs of corresponding points in the goal and similar figures (i.e., in $\langle S_1, R_1\rangle$ superimposed with the similar figure) to determine the point of similarity (P_s), and then, constructs a line through the point of similarity and a point on the similar figure which corresponds to a needed point of the goal figure. (Point of similarity rules are denoted r_{PS}.) Finally, the obtained point on the goal figure is used as a basis for constructing the goal square.

The second class is well represented by the problem, "Given angles B and C of a triangle, and the median M_a to side a, construct the triangle." The corresponding solution rules begin similarly by applying a similar figures rule (r_{gs}) to two given angles to construct an arbitrary sized triangle similar to the goal triangle, with medians, altitudes, etc., as required. Then a modified point of similarity rule (r_{PS}) is used to determine the point of similarity (P_s, the vertex of the non-given angle), and to construct the given segment (e.g., M_a), such that one endpoint of the segment is the point of similarity, and such that the segment coincides with the corresponding segment in the similar triangle. Finally, a line is constructed, through the other endpoint of the constructed segment parallel to the side of the similar triangle that is opposite to the point of similarity. The remaining sides of the goal triangle are obtained by extending two sides of the similar triangle to intersect the constructed parallel line.

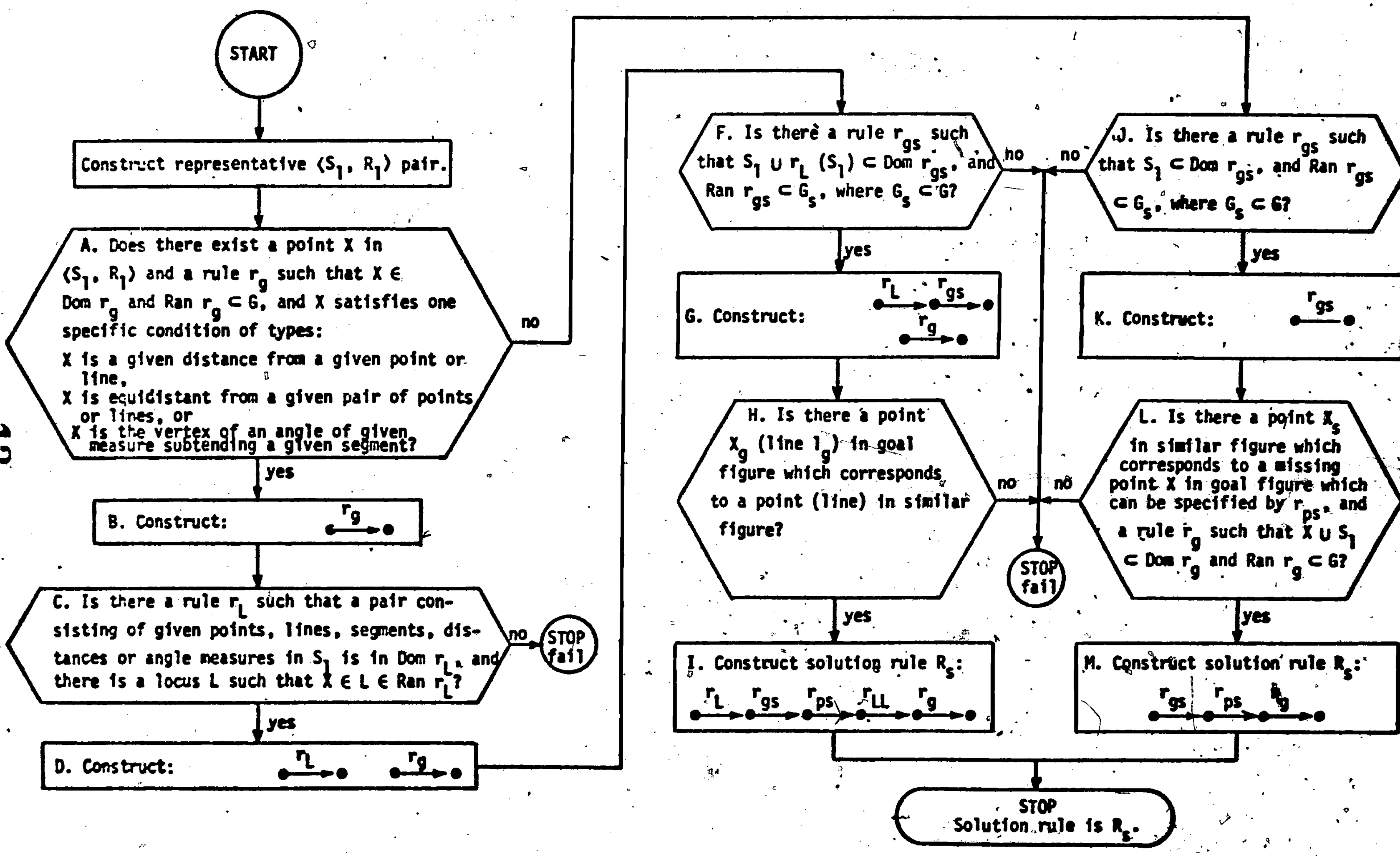
The solution rules for the third class of problems differ in that the first step in each is to use an r_L rule to construct a locus of points which contains a critical point, specifically the center of the goal circle. In the problem, "Given a line and two points (A and B) on the same side of the line, construct a circle tangent to the line which passes through the two given points," for example, the locus of points (L) equidistant from the two given points contains the center of the goal circle. Also, the point of similarity is the intersection of the locus and the given line. The second step is to construct a similar figure (circle, C_1), thich satisfies part of the goal condition. In our example, a circle is constructed with center on the constructed locus and tangent to the given line. Next, another version of the point of similarity rule is applied; this time the point of similarity (P_s) and a given point on the goal figure (e.g., B) are used to determine a corresponding point (B') on the <u>similar</u> circle. Then, parallel lines involving corresponding points are constructed to determine the center of the goal circle. Finally, the goal circle is actually constructed.

<u>The similar figures rule</u>

The higher order rule shown in Figure 4 (together with a set of applicable lower order rules) provides a sufficient basis for solving all of the sampled pattern of similar figures problems. Furthermore, the higher order rule appears to reflect the underlying semantics (Figure 4 found on the following page). For example, let us see how a solution rule for the first illustrative problem above (inscribing a square in a triangle) can be generated by application of the higher order rule. The first decision making capability (A) asks essentially whether a point X is needed to serve as the center for a goal circle. As the goal figure is a square, the answer is obviously no. Decision making capability J then asks if there is a goal similar figure rule (r_{gs}) which applies to representing stimulus S_1 and generates squares that satisfy part (G_s) of the goal condition (i.e., the range of r_{gs} is contained in G_s which in turn contains G - equivalently, anything which satisfies G, satisfies G_s, but not necessarily conversely). The lower order rule, "Construct a square in a triangle with one side coincident with one side of the triangle and one vertex on another side of the triangle," satisfies these conditions so the rule is retrained as indicated in operation K.

Decision making capability L asks two things: (1) Is there a point X_s which corresponds to a missing point X in the goal square? (2) Is there a rule r_g such that

Figure 4



START
Construct representative $\langle S_1, R_1 \rangle$ pair.
A. Does there exist a point X in $\langle S_1, R_1 \rangle$ and a rule r_g such that $X \in$ Dom r_g and Ran $r_g \subset G$, and X satisfies one specific condition of types:
X is a given distance from a given point or line,
X is equidistant from a given pair of points or lines, or
X is the vertex of an angle of given measure subtending a given segment?
yes
no
B. Construct: r_g
C. Is there a rule r_L such that a pair consisting of given points, lines, segments, distances or angle measures in S_1 is in Dom r_L, and there is a locus L such that $X \in L \in$ Ran r_L?
no
STOP fail
yes
D. Construct: r_L r_g
F. Is there a rule r_{gs} such that $S_1 \cup r_L(S_1) \subset$ Dom r_{gs}, and Ran $r_{gs} \subset G_s$, where $G_s \subset G$?
no
yes
G. Construct: r_L r_{gs} r_g
H. Is there a point X_g (line l_g) in goal figure which corresponds to a point (line) in similar figure?
no
yes
I. Construct solution rule R_s: r_L r_{gs} r_{ps} r_{LL} r_g
J. Is there a rule r_{gs} such that $S_1 \subset$ Dom r_{gs}, and Ran $r_{gs} \subset G_s$, where $G_s \subset G$?
no
yes
K. Construct: r_{gs}
L. Is there a point X_s in similar figure which corresponds to a missing point X in goal figure which can be specified by r_{ps}, and a rule r_g such that $X \cup S_1 \subset$ Dom r_g and Ran $r_g \subset G$?
no
yes
STOP fail
M. Construct solution rule R_s: r_{gs} r_{ps} r_g
STOP
Solution rule is R_s.

the stimulus S_1, supplemented with the point X (i.e., $X \cup S_1$), is in the domain of r_g, and r_g generates a goal-like figure (ran $r_g \subset G$)? In short, is there a point X_s in the similar square which corresponds to a point X from which the goal square may be constructed? Clearly, there is such a point X_s and the rule, "Determine the distance from a point to a given line segment and construct a square with sides of that length" satisfies the necessary conditions. Operation M forms the solution rule consisting of the two rules above with the point of similarity rule between them.

To see how the higher order rule works with the second class of problems, consider the second illustrative problem above (constructing a triangle, given two angles and a median). In this case, the answers to decision making capabilities A and J are again "no" and "yes," respectively. Here, r_{gs} is, "Construct a triangle of arbitrary size using two given angles and add parts corresponding to given segments." At decision making capability L there is a point X in the goal figure, the endpoint of median M_a, which can be specified by r_{PS}. Operation M again forms the solution rule.

Notice that the first two classes of problems involve the same path in the higher order rule. Each solution rule requires a goal similar figure rule (r_{gs}), the point of similarity rule (r_{PS}), and a goal constructing rule (r_g). The only difference is whether the goal and similar figures are squares or triangles, with all that implies for the particular r_{gs} and r_g rules required. In short, this example illustrates how what may appear initially to be basically different kinds of problems may turn out to have a common genesis.

The third problem (constructing a circle tangent to a given line and passing through two given points) illustrates the other path through the higher order rule. In this case, if we knew the center (X) of the desired circle, we could solve the task. Furthermore, this missing point X is on a locus, namely the locus of points equidistant from the two given points. Hence, decision making capabilities A and C are satisfied, and we retain the circle constructing rule (r_g) and the perpendicular bisector rule. Decision making capability F asks if there is a rule (r_{gs}) which applies to the stimulus S_1 as modified by the output of the locus rule (i.e., $S_1 \cup r_{PB}(S_1)$). Condition F is satisfied by a rule that generates circles with centers on a given line (the locus) and tangent to another given line. The answer to the decision making capability H is also "yes." The two given points on the goal figure obviously correspond to two points on the similar circle. By operation I, the solution rule follows directly: "Construct the locus of points equidistant from the two given points; construct a circle with center on that locus tangent to the given line; apply the point of similarity rule, and then the parallel line rule to determine the center of the goal circle; construct the goal circle using this center and the distance between it and a given point of radius."

It should be noted that in one of the sampled tasks the "locus" is given. The easiest way to handle this special case is to simply add an identity locus constructing rule as before. It would also be a simple matter to modify the higher order rule to take this possibility into account by asking, prior to or at decision making capability C, whether there is a line in S_1 which contains X.

Combined rule for two-loci and similar figures problems

It would appear from our analysis that the two higher order rules, together with the necessary lower order rules, would provide an adequate basis for solving the sampled two loci and similar figures problems and others like them. Indeed, there are two possible modes of solution in the case of one of the sampled similar figures tasks: "Inscribe a square in a right triangle so that two sides of the square lie on legs of the triangle, and one vertex of the square lies on the hypotenuse." Instead of using the pattern of similar figures, as illustrated in our first example, the pattern of two loci rule can be used to construct the bisector of the right angle. The intersection of this locus with the hypotenuse (the other locus) is the "missing point" X and provides a sufficient basis for constructing the goal square.

Although it is not always critical to distinguish between different modes of problem solving, any complete account designed to reflect human behavior must specify why one mode of solution is to be preferred over another (cf. Scandura, 1973a, Ch. 8). In the present case, there are two possible ways of handling this. First, we can add a higher order selection rule to the rule set which says simply, if both higher order rules apply, select the pattern of two loci. The rationale is that the pattern of two loci rule will generally yield a simpler method of solution.

A second way to handle the problem is to devise a single higher order rule which combines the advantages of both higher order rules. The higher order rules in Figures 3 and 4 can be combined to yield the higher order rule depicted in Figure 5. The path in this higher order rule designated ⟨1,2,3,4⟩ corresponds to that path of the two loci higher order rule which deals with those cases where the two loci may be found in either order. The path ⟨1,2,3,A,B,C,D,E,4⟩ deals with those two-loci problems where one locus must be found before the other. The other two paths correspond to the similar figures higher order rule.

PATTERN OF AUXILIARY FIGURES

Not all compass and straightedge problems can be solved via the pattern of two loci or the pattern of similar figures. In this section, we describe a higher order rule for dealing with the third class of problems identified by Polya (1962), the pattern of auxiliary figures. We also show how the combined higher order rule (above) may be extended to account for essentially all of the construction problems identified by Polya (1962).

Auxiliary figures higher order rule

Our initial analysis was based on a sample of five diverse auxiliary figures problems. One of the problems used was, "Given the three medians of a triangle, construct the triangle."

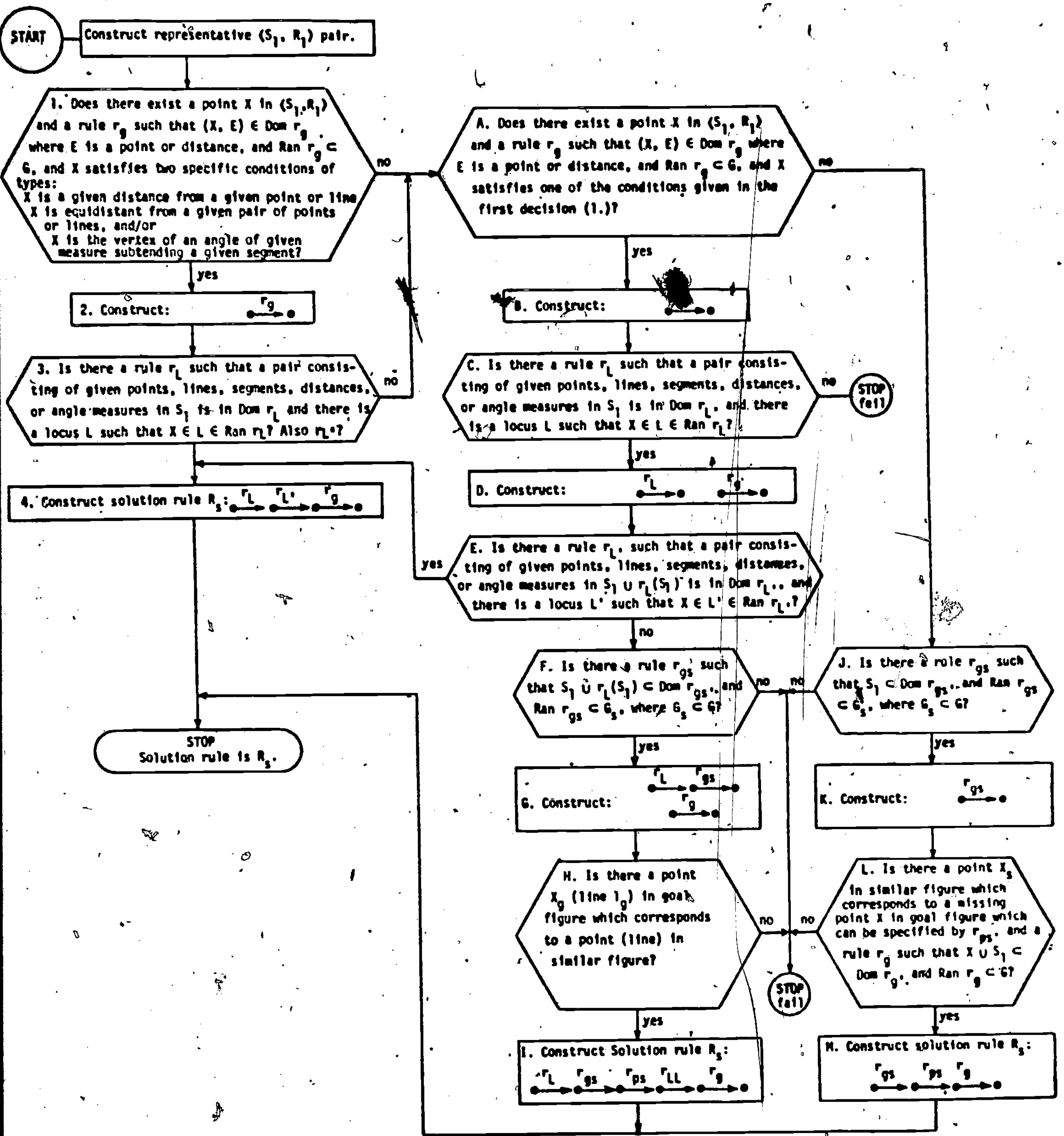
The analysis proceeded as before. First, we identified a procedure for solving each problem. Then, we looked for similarities among the solution rules and identified the component rules involved. In general, the required goal figures were not constructable via either the two loci or similar figures higher order rules. However, in each case the goal figure could be obtained from an (auxiliary) figure that was constructable from the given information. In the problem above, for example, a triangle can be constructed from segments one-third the lengths of the given medians. The goal figure is obtained by extending two of the sides of this auxiliary triangle to the respective median lengths and drawing lines through the resulting endpoints.

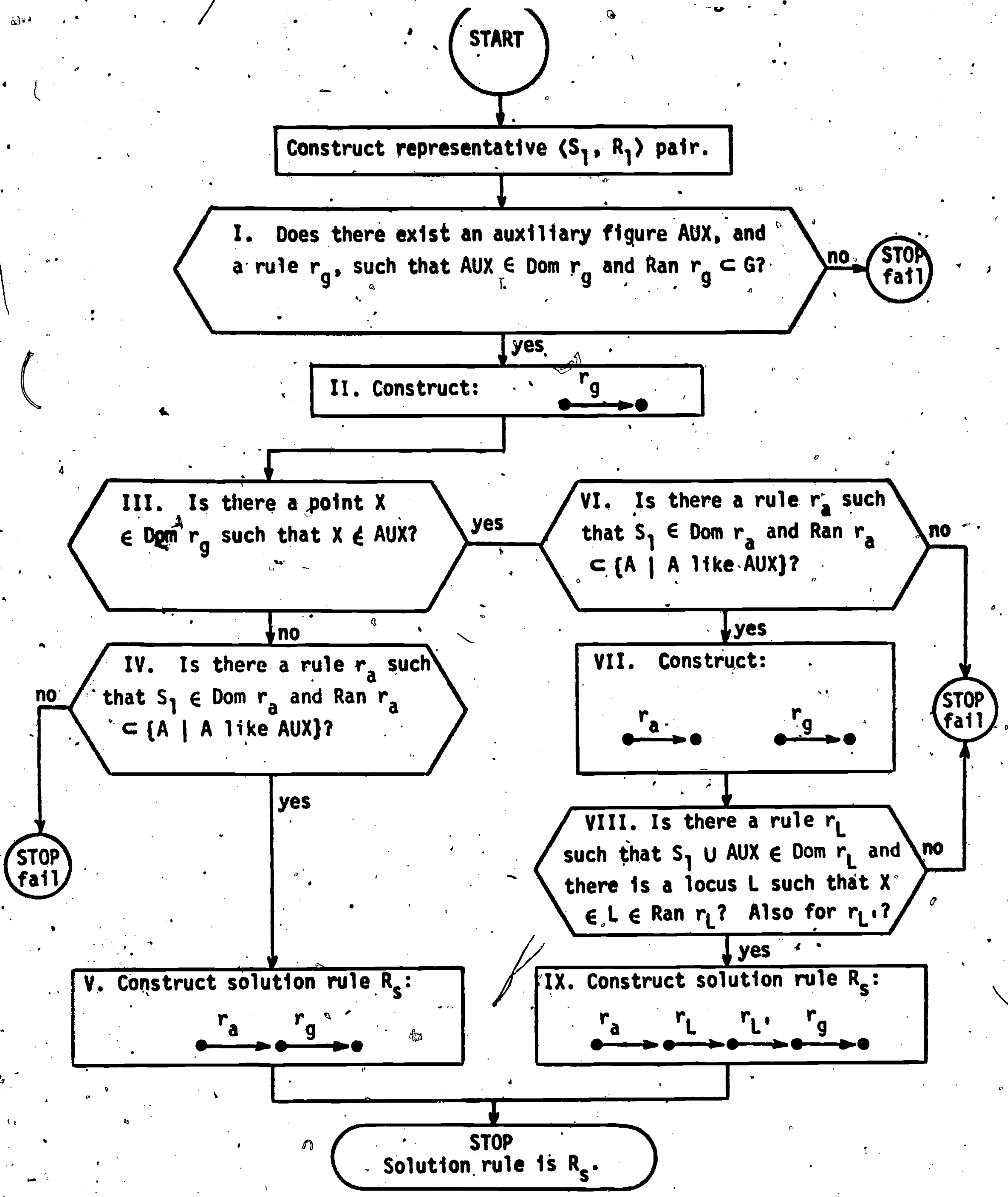
The analysis resulted in the auxiliary figures higher order rule shown in Figure 6. This higher order rule generates a solution rule for the illustrative task above as follows. First, an arbitrary representation for the solved problem $\langle S_1, R_1 \rangle$ is constructed. In this case, an arbitrary triangle is "sketched," and its medians are represented on it. The first decision asks whether there is (1) an auxiliary figure, and (2) a rule r_g which operates on the auxiliary figure and generates the goal figure. In this task, there is such an auxiliary figure, a triangle having sides one-third the lengths of the given medians.[13] In addition, the rule, "Extend the constructed segments to their given lengths and draw lines through their endpoints," satisfies condition (2). The next decision (III) asks whether or not a point is needed, in addition to the auxiliary figure, to construct the goal. Here, the answer is "no"; no other point is needed. Finally, decision IV asks if there is an auxiliary figure construction rule (r_a) available whose domain contains S_1 ($S_1 \in$ Dom r_a)

---

13. See Appendix F.

Figure 5


START
Construct representative (S1, R1) pair.
1. Does there exist a point X in (S1,R1) and a rule rg such that (X, E) ∈ Dom rg where E is a point or distance, and Ran rg ⊂ G, and X satisfies two specific conditions of types:
X is a given distance from a given point or line
X is equidistant from a given pair of points or lines, and/or
X is the vertex of an angle of given measure subtending a given segment?
no
yes
2. Construct:
rg
3. Is there a rule rL such that a pair consisting of given points, lines, segments, distances, or angle measures in S1 is in Dom rL and there is a locus L such that X ∈ L ∈ Ran rL? Also rL'?
no
4. Construct solution rule Rs:
rL
rL'
rg
STOP
Solution rule is Rs.
A. Does there exist a point X in (S1, R1) and a rule rg such that (X, E) ∈ Dom rg where E is a point or distance, and Ran rg ⊂ G, and X satisfies one of the conditions given in the first decision (1.)?
no
yes
B. Construct:
C. Is there a rule rL such that a pair consisting of given points, lines, segments, distances, or angle measures in S1 is in Dom rL, and there is a locus L such that X ∈ L ∈ Ran rL?
no
STOP
fail
yes
D. Construct:
rL
rg
yes
E. Is there a rule rL, such that a pair consisting of given points, lines, segments, distances, or angle measures in S1 ∪ rL(S1) is in Dom rL', and there is a locus L' such that X ∈ L' ∈ Ran rL'?
no
F. Is there a rule rgs such that S1 ∪ rL(S1) ⊂ Dom rgs, and Ran rgs ⊂ Gs, where Gs ⊂ G?
no
no
yes
G. Construct:
rL
rgs
rg
H. Is there a point Xg (line lg) in goal figure which corresponds to a point (line) in similar figure?
no
no
STOP
fail
yes
I. Construct Solution rule Rs:
rL
rgs
rps
rLL
rg
J. Is there a rule rgs such that S1 ⊂ Dom rgs, and Ran rgs ⊂ Gs, where Gs ⊂ G?
yes
K. Construct:
rgs
L. Is there a point Xs in similar figure which corresponds to a missing point X in goal figure which can be specified by rps, and a rule rg such that X ∪ S1 ⊂ Dom rg, and Ran rg ⊂ G?
yes
M. Construct solution rule Rs:
rgs
rps
rg


START
Construct representative $\langle S_1, R_1 \rangle$ pair.
I. Does there exist an auxiliary figure AUX, and a rule r_g, such that AUX $\in$ Dom r_g and Ran $r_g \subset G$?
no
STOP fail
yes
II. Construct: r_g
III. Is there a point X $\in$ Dom r_g such that X $\notin$ AUX?
yes
no
IV. Is there a rule r_a such that $S_1 \in$ Dom r_a and Ran $r_a \subset \{A \mid A \text{ like AUX}\}$?
no
STOP fail
yes
V. Construct solution rule R_s: r_a r_g
VI. Is there a rule r_a such that $S_1 \in$ Dom r_a and Ran $r_a \subset \{A \mid A \text{ like AUX}\}$?
no
STOP fail
yes
VII. Construct: r_a r_g
VIII. Is there a rule r_L such that $S_1 \cup$ AUX $\in$ Dom r_L and there is a locus L such that X $\in$ L $\in$ Ran r_L? Also for $r_{L'}$?
no
yes
IX. Construct solution rule R_s: r_a r_L $r_{L'}$ r_g
STOP
Solution rule is R_s.

and whose range contains the auxiliary figure (i.e., Ran $r_a \subset$ { A|A is like AUX}). In this case, the rule, "Construct a triangle from segments one-third the lengths of three given segments (medians)" satisfies these conditions and operation V constructs the solution rule, "Construct a triangle having sides one-third the length of the given medians; extend two segments of the constructed triangle to the respective median lengths, and draw lines through the endpoints of the medians to construct the goal triangle."

The other path through the higher order rule may be illustrated using the task, "Given the four sides a,b,c,d of a trapezoid (a<c), construct the trapezoid." Again, the answer to decision I is "yes." (Where the answer is "no," the higher order rule fails.) The triangle with c-a, b, d as sides, serves as the auxiliary goal figure and the goal rule, "Through corner points of an auxiliary figure and through another point not in the auxiliary figure, draw segments to complete the goal," is selected. Unlike the first path, however, the answer to decision III is "yes" since the goal rule (r_g) acts on pairs (X ∪ AUX) consisting of an auxiliary figure and a critical point X. The next decision (IV) asks if there is a rule r_a that constructs the auxiliary figure from given information. This condition is satisfied by the r_a rule which constructs the auxiliary triangle from the sides of a trapezoid. Decision VIII asks whether there are two locus rules (r_L and $r_{L'}$) which apply to the auxiliary figure and/or other given information (S_1) and whose ranges contain X. The circle rule (r_C), applied to different portions of S_1∪ AUX, plays the role of both locus rules. The solution rule (Operation IX) is a concatenation of the component rules.

Combined two-loci, similar and auxiliary figures higher order rule

Taken collectively, the three higher order rules described above can be used to construct solution procedures for a wide range of geometry construction problems. Furthermore, they appear compatible both with human behavior and with the heuristics originally identified by Polya (1962).

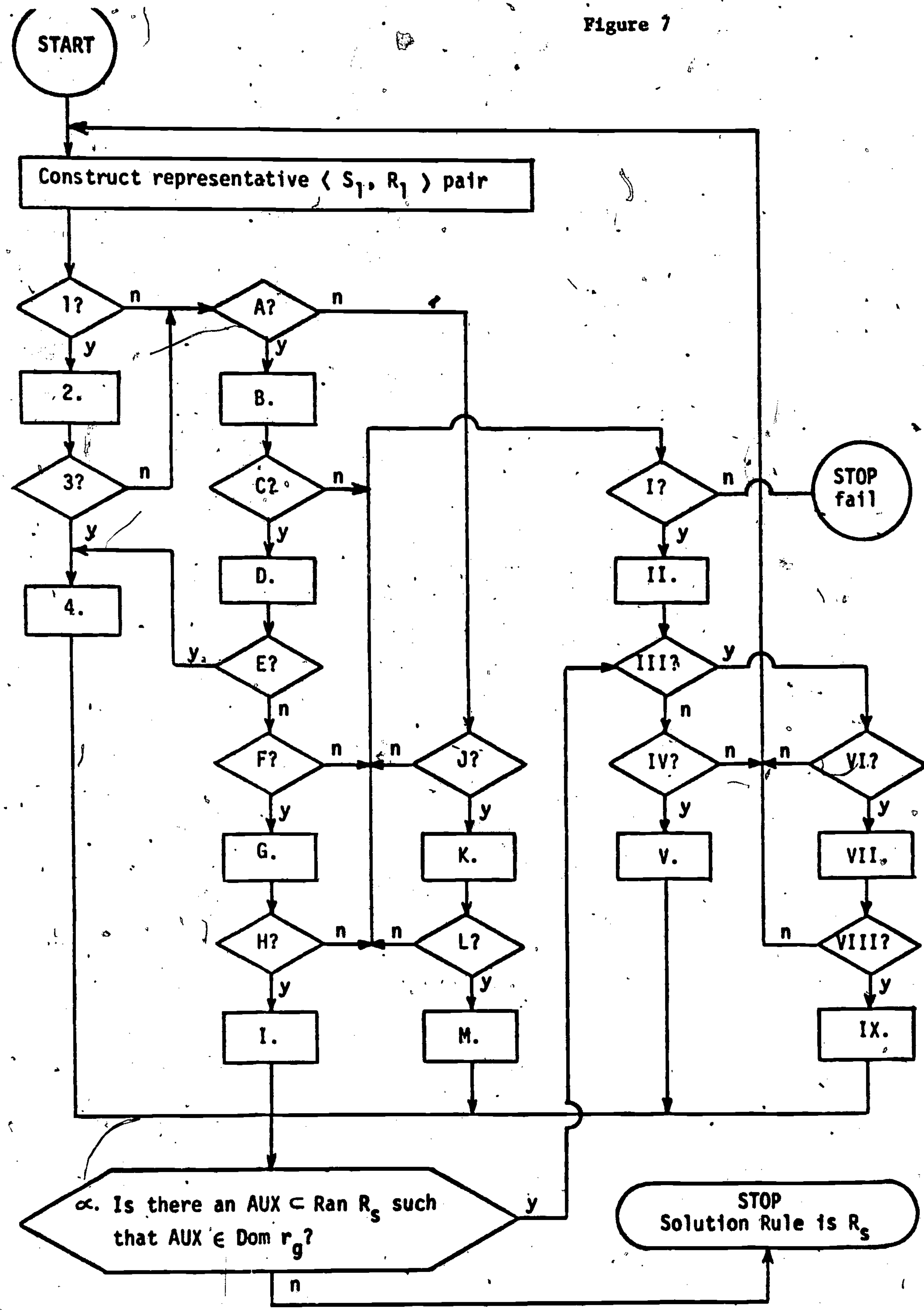
This is not meant to imply, however, that the three higher order rules are unrelated to one another. Both the needed point X in the pattern of two loci, and the similar figure in the pattern of similar figures can be regarded as special auxiliary figures. Indeed, one could modify the auxiliary figure higher order rule so that it, together with the relevant lower order rules, would account for all three classes of problems. In addition, the similar and auxiliary figures higher order rules may be viewed as progressive generalizations of the two-loci higher order rule. It is not difficult to conceive of third level higher order generalization rules which have the two loci higher order rule and a similar or auxiliary figure as inputs, and a more general higher order rule in which a similar or auxiliary figure is substituted for the missing point X, as the corresponding output.

Alternatively, the combined two-loci, similar figures higher order rule (Figure 4) can be extended to include auxiliary figures. In addition, the extended higher order rule depicted in Figure 7 allows recursion on the higher order rules.

To see this, notice that the higher order rule shown in Figure 6 can terminate at several points without finding a solution rule. In some problems this is unavoidable; there may not be an auxiliary figure from which the goal figure can be constructed. Sometimes, however, there is an auxiliary figure, but one which is not directly constructable from the given information. Such auxiliary figures can often be constructed via the pattern of two-loci, the pattern of similar figures, or the pattern of auxiliary figures itself. In those cases where such an auxiliary figure exists, we allow for this possibility by returning control to the start of the combined higher order rule in order to derive an r_a rule for constructing the auxiliary figure. Once an auxiliary figure (r_a) rule has been derived, the original procedure resumes.

To see how this higher order rule works, consider the following task, "Construct a trapezoid given the shorter base a, the base angles A and D, and the altitude H_t." As

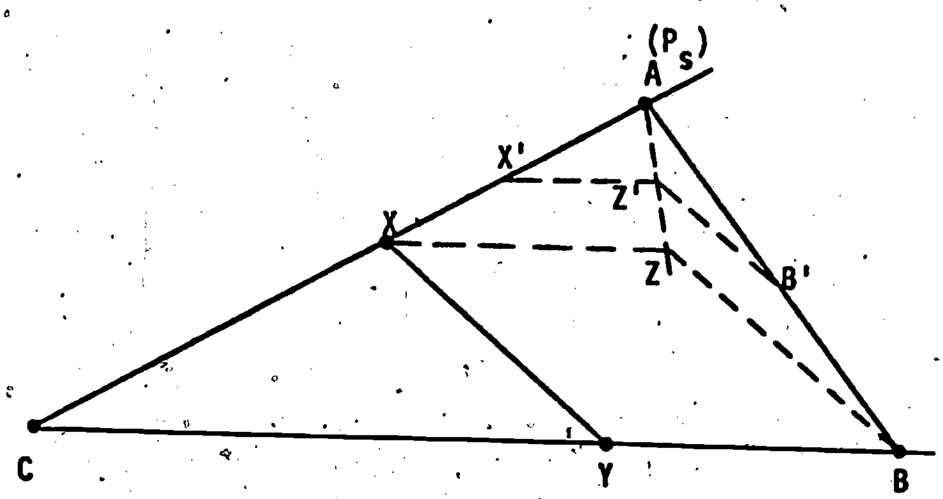
Figure 7


START
Construct representative ⟨ S_1, R_1 ⟩ pair
1?
2.
3?
4.
A?
B.
C?
D.
E?
F?
G.
H?
I.
J?
K.
L?
M.
I?
II.
III?
IV?
V.
VI?
VII.
VIII?
IX.
STOP fail
α. Is there an AUX ⊂ Ran R_s such that AUX ∈ Dom r_g?
STOP Solution Rule is R_s
y
n

in the trapezoid example given earlier, the needed auxiliary figure is the triangle having sides c-a, b, and d. But, this triangle is not directly constructable from the given information. None of the assumed lower order rules is adequate, so the higher order rule breaks down at step VI. The flow of control therefore returns to step 1 with the aim of constructing the auxiliary figure.[14] Beginning here, the problem of constructing this auxiliary figure is a straightforward similar figures task, one in fact which we had sampled.

The higher order rule of Figure 7 also generates solution rules for even more complex problems, provided we assume the necessary component rules. For example, consider the problem, "Given three noncollinear points A, B, and C, construct a line XY which intersects segment $\overline{AC}$ in the point X and segment $\overline{BC}$ in the point Y, such that segments $\overline{AX}$, $\overline{XY}$, and $\overline{YB}$ are all of the same length."

Figure 8



The reader may wish to derive the solution rule for this more difficult problem himself. (Hint: Several recursions are required. For details see the unabridged report.)

DISCUSSION

Summary

In summary, a quasi-systematic method for characterizing heuristics involved in problem solving was proposed and illustrated with compass and straightedge constructions in geometry. Higher order rules, together with corresponding sets of lower order rules, were constructed for the two-loci, similar figures and auxiliary figures problems identified by Polya (1962). First, the two-loci heuristic of Polya was made precise. We saw how decision making capabilities (decisions), and particularly the conditions used to define decisions, play a central role in higher order rules. The similar figures and auxiliary figures heuristics were similarly formulated. We also showed how the two-loci and similar figures higher order rules could be combined to form one higher order rule, which (together with appropriate lower order rules) provides a basis for solving both

14. This involves memory and is not indicated in the flow diagram.

kinds of problems. Finally, a combined two-loci, similar and/or auxiliary figures higher order rule was constructed. This higher order rule allows recursive returns to components of the higher order rule, corresponding to the individual higher order rules, and was considerably more powerful than the others. Its use on some complex problems was illustrated.

Overall, the analyses demonstrated the viability of the analytic method. The higher order rules identified were precise, compatible with the heuristics identified by Polya, and intuitively seemed to reflect the kinds of relevant knowledge that successful problem solvers might have.

The central role played by semantics in the analysis should be emphasized. The meaning of each task was represented by a goal figure $\langle S_1, R_1 \rangle$ representing the given goal situation $\langle S_0, R_0 \rangle$. The relations among, and properties of, the elements of these figures, together with the domains and ranges of individual rules, were reflected directly in the higher order rules. Although little attention was given to the formal representation of semantic features, the goal figures clearly placed powerful constraints on the rules selected at each stage in applying the higher order rules. Representation in terms of some arbitrary (e.g., random) syntax, unconstrained by goal figures, would have necessitated backup capabilities and, in principle, could easily increase the number of possible construction rules at each stage beyond any reasonable computational capability. That is, without the constraints imposed by the goal figures, the number of possible points, arcs, and lines that might be constructed could be almost unlimited. The effect of using goal figures is very much the same as that referred to by Winston (1972) in a recent paper on vision. He argued that although the number of combinatorially possible arrangements of vertex types (Guzman, 1968) is very large, the number of types that yield real figures is much smaller.

Limitations

Nonetheless, the present study has certain limitations which, in principle, could be overcome. First, as in existing state space formulations, all of the higher order operations were limited to compositions of rules. In future research, more attention should be given to other kinds of operations. Generalization, restriction, and selection rules (e.g., Scandura, 1973a), for example, might well be expected to play an important role in problem solving.

There are a variety of ways in which such rules might enter. (a) In discussing the two-loci higher order rule, we have already seen how the scope of a decision (making capability) may be generalized to generate solution rules for a broader range of problems. In particular, we saw how the first decision, which was initially restricted to situations where the desired point X was a given distance from two given points, could be generalized, for example, to allow the point to be the same distance from two given points. It is not hard to envisage a generalization rule by which such shifts might be made. The relationships observed previously between the missing points X, and the similar and auxiliary figures, suggest another kind of generalization involving the identified higher order rules.

(b) There are a wide variety of construction problems which might require the independent derivation of more than one missing point X, similar figure, or auxiliary figure. As a simple example, consider the task of constructing two circles, one of which is to be inscribed in a given triangle and the other, to pass through its vertices (i.e., to circumscribe the triangle). In this case, the problem can be solved by applying the two-loci higher order rule twice. The higher order derivation rule here can be thought of as a generalization of the two-loci rule in which two or more applications (i.e., recursions) may be allowed. One can easily conceive of a simple higher order generalization rule which operates on rules and generates corresponding rules which are recursive. The combined two-loci, similar and auxiliary figures higher order rule is one possible consequence of apply some such higher order rule.

(c) If we had allowed unsolvable variants of the problems considered, truly viable solution rules would have to be appropriately restricted. The solution rule for "constructing a triangle with sides of predetermined length," for example, works only when the sum of each pair of sides of the triangle is greater than the third. A completely adequate solution rule would have to test this possibility. It is possible to conceive of higher order rules, which operate on rules of various kinds together with special restrictions (e.g., the triangle inequality) to generate correspondingly restricted rules.

(d) It is also possible to conceive of three dimensional analogues of compass and straightedge constructions. In this case, the higher order rules would operate on the usual two dimensional construction rules and would generate their three dimensional analogues. For example, a rule for constructing the locus of points equidistant from a given line (i.e., a pair of lines) corresponds to a three dimensional rule which constructs a cylinder about the line.[15]

A second limitation is that nowhere did deduction play a role in our analysis. In solving constructions, real people frequently attempt to justify logically the various constructions they make. Constructing a triangle given its three medians, for example, requires that a person know or deduce the fact that the medians intersect at a point two-thirds of the way from each vertex to the opposite midpoint (see footnote 13). To this extent, our analysis is limited and may not adequately reflect human knowledge. Our rules reflect semantics, but not inference. Extension of the proposed analysis to deduction should be a first order of business. It is likely that existing geometry theorem proving systems (e.g., Gelernter, 1959) may be useful in this regard.

A third major limitation of this research is that cumulative effects of learning were not considered: each problem in our analysis was considered as de novo. If one wishes to characterize solutions to problems in a given class (e.g., the two-loci tasks) relative to a fixed, self-sufficient set of rules, some fairly complex rules (e.g., the angle vertices rule) must be included. Furthermore, and in many ways more important, such characterizations, at any particular level of analysis in a task domain tend to lack flexibility. The atomic elements are so large, relatively speaking, that there are many intermediate level problems that cannot readily be solved using such rule sets exclusively. Also important from the standpoint of behavioral analysis, it is doubtful that such lower order rules would adequately reflect the knowledge had by most subjects assumed to know the identified higher order rules. Such subjects would almost certainly also know a wide variety of simpler construction rules, even though we might not explicitly include them in a rule set determined by sampling complex problems of the sort we used. Future work is planned which is designed to meet many of these objections.

Future directions

The method of analysis used in the present research is based on Scandura's (1973a) theory of structural learning, more particularly on those aspects of it which deal with competence. The aim of the latter is to specify (hopefully mechanizable) procedures which characterize the knowledge underlying given classes of behaviors (e.g., problem solutions) that one might wish to attribute to an idealized knower. As noted, our approach to this problem involves the invention of finite sets of rules (including higher order rules which may operate on other rules as well as on data elements) which can be applied as indicated for example, to generate problem solutions.

This level of theory, of course, applies only at an analytic level in the sense generative grammars account for language behavior. The relevance of the theory to actual human behavior, or, for that matter, to the design of artificial intelligence systems, depends fundamentally on our ability to specify mechanisms by which such rules are to interact in specific situations, and what effect if any such interaction has on the nature of the rule set itself.

---

15. Implicit in the above examples is another limitation to which we have indirectly referred previously. Our original analyses were limited almost exclusively to single higher order rules. In no case did we attempt to identify rules which may operate on higher order rules, although our examples make it clear that we could have done so. The problems involved in accomplishing this would be practical rather than theoretical.

The structural learning theory (Scandura, 1973a) is partly concerned with the specification of such mechanisms. The theory rests on the fundamental and widely held assumption that in problem solving people are attempting to achieve some goal. In the simplified version of the theory considered here, the basis mechanism which governs the use of available rules is as follows: (A) The subject tests his available rules (r) to see if one (or more) of them satisfies the given goal situation (i.e., If $S_0 \in$ Dom r and Ran r $\subset$ Goal). If so, the subject will apply it. (B) If a subject does not have a rule available for achieving a given goal, then control automatically shifts to the higher level goal of deriving a procedure which will satisfy the original goal. (C) If a higher level goal has been satisfied (that is, if some new rule has been derived which contains the stimulus situation in its domain and whose outputs satisfy the original goal criterion), the derived rule is added to the set of available rules and control reverts back to the previous goal. The third hypothesis allows control to return to lower level goals once a higher level goal has been satisfied. (For more general and rigorously formulated sets of hypotheses see Scandura, 1973a.)

Putting all this together, we see that if an appropriate higher order rule is available when control shifts to a higher level goal, then the higher order rule will be applied and control will automatically revert to the original goal. The subject will then apply the newly derived rule and solve the problem. If the subject does not have a higher order rule available for deriving a procedure that works, then control is presumed to move to still higher levels (e.g., deriving a rule for deriving a rule that works). Although this process is assumed to go on indefinitely in the idealized theory, memory places strict limits in actual applications.

Even this simple assumption provides an adequate basis for generating predictions in a wide variety of problem solving situations. Consider the problem of converting a given number of yards into inches. There are two possible ways in which a subject might solve the problem. The first is to simply know, and have available, a rule for converting yards directly into inches: "Multiply the number of yards by 36." In this case, the subject need only apply the rule according to hypothesis (A). The other way is more interesting, and involves the entire mechanism as described above. Here, we assume that the subject has mastered one rule for converting yards into feet, and another for converting feet into inches. The subject is also assumed to have mastered a higher order composition rule.

In the second situation the subject does not have an applicable rule which is immediately available, and, hence, according to hypothesis (B), he automatically adopts the higher level goal of deriving such a procedure. Then, according to the simple performance hypothesis (A), the subject applies the higher order composition rule to the rules for converting yards into feet and feet into inches. This yields a new composite rule for converting yards into inches. Next, control reverts to the original goal by hypothesis (C) and, finally, the subject applies the newly derived composite rule by hypothesis (A) to generate the desired response.

Moreover, this mechanism provides a basis for an efficient characterization of learning, since, according to hypothesis C, newly derived rules are added to the knowledge base (rule set). Such (additional) rules are in no way distinguished from any others in the rule set; for example, they may serve as component rules in new higher order rule applications. (Also, it should be noted that derived rules may themselves be of higher order and may, thus, be used to satisfy future higher level goals.)

To see how knowledge may cummulate according to this mechanism, let us assume that the learner initially knows rules for converting miles into yards, yards to feet, feet to inches, and the higher order composition rule above. Suppose also that the learner is first presented with the problem of converting miles to inches. In this situation, the learner will fail to solve the problem, since the composition rule we specified above applies only to pairs of rules. (We assume that it does not apply to itself.) However, if the problem of converting yards to inches is presented first, the subject will solve it as before, and derive a yards to inches rule in the process. Further, if the miles-to-inches problem is then presented, it can be solved using the derived yards to inches rule and the miles to yards rule as components. Although this example is obviously very simple, it does illustrate the potential importance of problem sequence in a growing (learning) system.

Although other investigators have made use of similar notions in varying degrees, the type of mechanism proposed appears to make more general use of rule and higher order rule constructs. Frequently, for example, procedures which are allowed to operate on procedures are not themselves part of the knowledge base; they are viewed as control processes. (In the present case, only the learning mechanism itself acts as a control process.) Nor are newly derived solution procedures often added to the set of available procedures. Newell & Simon (1972, p. 135), for example, allow the Logic Theorist to add proved theorems to an initial set of axioms, but this is essentially at the level of data, upon which proof generation procedures operate, and not at the level of the procedures themselves. Viewing learning as "debugging" (e.g., Minsky & Papert, 1972) or as "means-ends" analysis (Newell & Simon, 1972) is essentially analogous to the introduction of higher order rules except that in these cases implicit restrictions are imposed on the allowable higher order rules.

In any case, most investigations in artificial intelligence have involved some kind of state space representation (e.g., Nilsson, 1971), with problem solving involving some type of search. No generally agreed upon way of representing learning seems to have emerged, however. Sometimes, learning is treated as the modification of parameters in evaluation functions which select 'promising' nodes for expansion (e.g., Samuel, 1959). In other cases, learning systems have been devised to reflect stimulus-response principles in psychology (e.g., Feigenbaum, 1961, Bower, 1972). Where considered by information processing psychologists who have adopted this point of view (e.g., Rumelhart, Lindsay, & Norman, 1972), learning involves the transformation of one state space to another (Scandura, 1973b).

Though the proposed representation may be formally equivalent, it is our belief, based on a variety of studies with human subjects (e.g., Scandura, 1973a), that it is not psychologically equivalent. For one thing, our search for basic psychological mechanisms (e.g., of learning), which reflect commonalities in human behavior, differs in important ways from that in computer simulation, where the essential goal is to parallel overt human behavior in complex instances of problem solving and where the basic mechanisms (e.g., means-ends analysis), therefore, are often judged on more immediately pragmatic grounds.

Irrespective of one's opinion on the issue, the laws which govern the interactions among individual rules are assumed to be fixed once and for all and have potentially important implications for computer implementation. In particular, the fixed mode of interaction would make it possible in principle to modify and/or extend an artificial intelligence system rule by rule, without having to worry about the effects of these changes on other parts. (This latter property appears to some extent to be shared by Newell and Simon's (1972) production systems.)

One of the major complications in current artificial intelligence research is that even minor changes in one part of a system may have unpredictable effects which may require compensating changes elsewhere. The switch to heterarchical systems (e.g., Minsky and Papert, 1972) in which control may shift among individual programs in some predetermined manner, does not appear to alleviate this problem.[16] In contrast to the above mechanism, the mode of control in heterarchical systems may vary from system to system, and worse, from the standpoint of debugging, may interact with the individual programs themselves. In short, the important point for artificial intelligence research is the possible advantage for implementation of a fixed mode of interaction.

Whether or not the mode of interaction is restricted to that proposed here is not the most crucial point. To the extent that artificial intelligence research may

---

16. See Appendix G.

benefit by taking account of such mechanisms, psychological research aimed at discovering what these mechanisms are would appear to be a first order of business for those interested in human thought. (For a "richer" theoretical mechanism which incorporates memory, see Scandura, 1973a, Ch. 10.)

With the foregoing in mind, an alternative which we are now pursuing is to begin initially with rule sets composed of simpler rules, and to allow these rule sets to grow gradually by interacting with a problem environment.[17] In the present case, only three atomic operators (lower order rules) will be introduced initially: (a) setting a compass to a given radius, (b) drawing a straight line (segment), and (c) using a set compass to make a circle. It is not immediately clear what the higher order rules should be but, presumably, any reasonably satisfactory rule set would include some types of simple composition, conjunction, and generalization higher order rules, together, possibly, with variants of the two loci and other higher order rules identified above. It should be emphasized in this regard that the initial selection of rules would not in itself be sufficient; the choice and sequencing of to-be-solved problems may also be expected to have important effects on both the rate and type of knowledge acquisition. For obvious reasons, computer implementation seems almost essential in this research and is the course we are pursuing.

IMPLICATIONS

Artificial Intelligence

The present research appears to have three general implications for work in simulation and artificial intelligence.

First, the rules we have identified may be implemented relatively easily (some have already been). As such, they would be useful either directly in systems concerned with geometric figures and constructions, or indirectly in research having more encompassing aims as described above.

Second, the results are suggestive of how the construction of at least certain artificial intelligence systems might be partially systematized. In this regard, the topic of compass and straightedge constructions is not nearly as important as is the fact that the analysis serves as a prototype for the proposed method of analysis. At the present time this method is being used to analyze the proofs contained in an experimental algebra I high school text based on axiomatics.

Third, our use of flow diagramming as a mode of representation of individual rules suggests that perhaps such representation might play a somewhat larger role in the exposition of future artificial intelligence research. The routine use of a large number of different and highly technical programming languages is often enough to turn away outsiders (such as ourselves) who might otherwise be interested.[18] The limitations of flow diagrams with regard to memory considerations may be a small price to pay for a more neutral and familiar form of representation. Furthermore, flow diagrams have a flexibility as to level of representation which is not shared by particular programming languages. This makes it possible to more readily represent basic components at a level of atomicity tailored to immediate needs, and to psychological reality (cf. Scandura, 1973a), rather than to basic components determined by some programming language. These comments, of course, apply only to psychological and expository considerations and say nothing of the more strictly technical problems of representation which must be dealt with in computer implementations.

Education

The results of this study also have both long range and immediate implications for education. The promising nature of the results attests to the practicability of the proposed approach as a means of identifying the knowledge underlying reasonably complex kinds of problem solving. In addition to serving as a prototype, the identified rules themselves could be helpful in teaching high school students how to solve compass and

17. See Appendix H.

18. See Appendix I.

straightedge construction problems.

By identifying precisely what it is that students must know (i.e., one possible knowledge base), these rules provide an explicit basis for both diagnosis and instruction. In particular, the methods of analysis formalized by Scandura (1973a) and developed empirically by Scandura and Durnin (1973) and Durnin and Scandura (1973) can be applied directly to assess the behavior potential of individual subjects on the individual rules, including the higher order ones. Operationalizing the knowledge of individual subjects in this way, and comparing this knowledge with the initial competence theory (i.e., set of rules), provides an explicit basis for remedial instruction (Durnin & Scandura, 1973). In effect, each subject can be taught precisely those portions of each competence rule which testing indicates he has not mastered.

Care was taken to help insure that the higher order rules reflect the kinds of ability individual subjects might have, or use. To the extent that the identified higher order rules are unknown to high school students, instruction in these rules ought to facilitate problem solving performance. The diagnostic and instructional efficacy of these higher order rules has been demonstrated in a recent field test (Scandura, Wulfeck, Durnin, & Ehrenpreis, 1974).

The above discussion of how knowledge is acquired through interaction of the learner with a problem environment also has educational relevance. Specifically, by assigning values to various objectives and costs to particular kinds of instruction (or rules), it should be possible to study the problem of instructional sequencing and optimization in a way which is both precise and relevant to meaningful education. We view this as a critically important problem for future research.

REFERENCES

Amarel, S. On representations of problems of reasoning about actions. In D. Michie (Ed.), Machine intelligence 3. New York: American Elsevier Pub., 1968, pp. 131-171.

Bower, G. A selective review of organizational factors in memory. In E. Tulving and W. Donaldson (Eds.). Organization of Memory. N.Y.: Academic Press, 1972.

Chomsky, N. Syntactic structure. The Hague: Mouton Press, 1957.

Durnin, J. H. & Scandura, J. M. An algorithmic approach to assessing behavior potential: Comparison with item forms and hierarchical technologies. Journal of Educational Psychology, 1973, 65, 262-272.

Ehrenpreis, W. and Scandura, J. M. An algorithmic base for curriculum construction: A field test in Mathematics. Journal of Educational Psychology, 1974, in press.

Feigenbaum, E. A. The simulation of verbal learning behavior. Proceedings of the Western Joint Computer Conference, 1961, 121-132.

Gelernter, H. Realization of a Geometry Theorem Proving Machine. Proc. Intern. Conf. Inform. Proc. Paris: UNESCO House, 1959. Reprinted in E. Feigenbaum & J. Feldman (Eds.) Computers and Thought. New York: McGraw-Hill Book Co., 1963.

Guzman, A. Decomposition of a visual scene into three-dimensional bodies. Proceedings of the Fall Joint Computer Conference, 1968, pp. 291-304.

Minsky, M. and Papert, S. Research at the laboratory in vision, language and other problems of intelligence. Artificial Intelligence Progress Report, Memo 252 Cambridge, Mass.: M.I.T. Aritificial Intelligence Laboratory, 1972.

Newell A., & Simon H. A. Human Problem Solving. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Nilsson, N.J. Problem-solving methods in artificial intelligence. New York: McGraw-Hill, 1971.

Polya, G. Mathematical discovery, Vol. 1. New York: Wiley, 1962.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long term memory. In E. Tulving and W. Donaldson (Eds.). Organization of Memory. New York: Academic Press, 1972.

Samuel, A. Some studies in machine learning using the game of checkers, IBM Journal of Research and Development. 1959, 3 (3), 211-229.

Scandura, J. M. Deterministic theorizing in structural learning: Three levels of empiricism. Journal of Structural Learning, 1971, 3(1), 21-53.

Scandura, J. M. Structural learning I: Theory and research. New York: Gordon & Breach, 1973. (a)

Scandura, J. M. The structure of memory: Fixed or flexible? University of Pennsylvania Structural Learning Series, Report 73, May, 1973. (b)

Scandura, J. M. & Durnin, J. H. Assessing behavior potential: Adequacy of basic theoretical assumptions. Journal of Structural Learning, 1973, in press.

Scandura, J. M., Wulfeck, W. H. II, Durnin, J. H., & Ehrenpreis, W. Diagnosis and instruction of higher order rules for solving geometry construction problems. University of Pennsylvania Structural Learning Series, Report 80, 1974.

Winston, P. H. Summary of selected vision topics. M.I.T. Artificial Intelligence Laboratory, Flash 30, July, 1972.

Appendix A.--The only really adequate way of determining whether a rule is compatible with human behavior is to effect a behavioral test; that is, to see whether a rule provides an adequate basis for assessing the behavior potential of individual subjects, thereby making it possible to predict the behavior of individual subjects on new instances (of the rule). (The theoretical foundations for such tests have been worked out and tested empirically [Scandura, 1971, 1973a, Scandura & Durnin, 1973; Durnin & Scandura, 1973].) The basic idea is to determine each subject's behavior potential with respect to each rule in an identified rule set, and then to use the theory as a basis for making predictions concerning performance on problems which require interactions among the rules. The closeness of fit between the predictions and observed behavior would provide a direct test of the adequacy of the rule set. A study reported in Scandura (1973a) on rule generalization was of this type. Since this was impractical in the present study, we adopted the weaker and less rigorous criterion of requiring that the rule sets be compatible with our intuition (cf. Chomsky, 1957).

Appendix B.--Strictly speaking, human subjects are presented with statements of problems as stimuli. Throughout this and our subsequent analyses we assume that the subject's initial subgoal is to interpret the goal statement (i.e., determine its meaning). The second subgoal is to solve the problem. In effect, the initial goal is divided into a pair of subgoals to be achieved in order. Our analysis is limited to the second part of this task, and then only on the assumption that there is no further division of the problem into subgoals. We also assume that the given problem statements can be uniformly and correctly interpreted.

Although we do not pursue the question here, we have reason to believe that forming subgoals is closely related to the question of (problem) representation (cf. Amarel, 1968).

Appendix C.--Other representations would probably be more efficient for computer implementation, since graphic systems are relatively complex to implement. For example, some sort of naming system for points, lines, etc. could be devised together with appropriate interpretive routines to identify relations of interest among elements. In fact, the naming system for triangles in common use, evolved for just this purpose; names for sides, vertices, medians, etc., if correctly interpreted, carry much information about relative position, intersections, etc.

Appendix D.--In the structural learning theory (Scandura, 1973a), it is assumed that the problem solver automatically tests the solution rule R_S to see if it satisfies a higher level goal condition. That is, is $S_O \in \text{Dom } R_S$ and $\text{Ran } R_S \subset G$? If the higher level goal is satisfied, control is assumed to revert to the original goal so that R_S will be applied.

Appendix E.--In evaluating alternative rule-based accounts for a given class of tasks, decisions must always be made concerning exactly how the computational load should be apportioned to the higher and lower order rules. Any number of alternatives exist; at one extreme, the lower order rules may do all of the computation, in which case a separate rule would be needed for each type of problem, and, at the other extreme, the component lower order rules may be of minimal complexity with the higher order rule assuming most of the computational burden. The requirement of compatibility with human knowledge, of course, substantially reduces the number of plausible characterizations.

Appendix F.--We do not attempt to spell out the procedures necessary for finding auxiliary figures. However, in all of the sampled auxiliary figures problems, it was necessary to construct a line parallel to some "distinguished" line through some "distinguished" point not on that line. Such procedures also frequently require special knowledge -- for example, that medians intersect at a common point that is 2/3 of the distance from the respective vertices to the midpoint of opposite sides. Such knowledge is frequently logically deducible, but for our purposes, may be represented in terms of simple "associations" for example, between triangles with their medians and the common intersection property.

Appendix G.--Scandura's (1973b) comments regarding relationships between the structural learning mechanism, and the notion of heterarchical control in systems of artificial intelligence (Minsky & Papert, 1972) may be relevant here.

"For a time artificial intelligence systems were viewed as wholes, as frequent complex programs. As work in the area progressed, the difficulties of building upon earlier work became increasingly clear because of the close interrelationships among various parts of such systems. To overcome this limitation, heterarchical, or modular planning has been used (e.g., Minsky & Papert, 1972). Heterarchical systems consist of sets of programs (modules) pertaining to syntax, semantics, line detection, and so on, together with a heterarchical executive which switches control among these "modules in accordance with a predetermined plan.

"Modules in heterarchical systems correspond, essentially to rules in the structural learning theory; the executive control structure corresponds to the basic mechanism. There is, however, an important difference between the two. In heterarchical systems, the basic goal is pragmatic. Such systems make it easier to modify and build upon previous work. No one seriously means to imply that heterarchical control reflects the way people perform, although in developing artificial intelligence systems intuitive judgements are sometimes made with this in mind.

"In contrast, the structural learning mechanism is assumed to be built into people (presumably from birth); it is not learned and need not be taught. While the rules a person knows may increase from time to time, the mechanism is assumed to remain constant.

"This is a strong claim, something which no responsible person would make concerning executive systems currently used in heterarchical systems. Among other things, it is very unlikely that an existing control system would be useful in systems other than the one for which it was designed. It is my contention that benefits might accrue in artificial intelligence and, of course, in simulation if structural learning like control structures were used [pp. 42-43]."

Appendix H.--Such rule sets have been called innate bases (Scandura, 1973a, Ch. 5). In general, innate bases lack the immediate, direct computing power of comparable rule sets composed of more complex rules but, theoretically at least, can grow to become more powerful.

Appendix I.--We realize, of course, that some computer specialists may not take our suggestion very seriously. We, however, find the work in simulation and AI highly suggestive for our own studies and hope in the interest of interdisciplinary communication that some readers may be moved more in this direction.