Tuning of Search of the Problem Space for Geometry Proofs*

John R. Anderson
Department of Psychology
Carnegie-Mellon University
Pittsburgh, Pa. 15213

To be reviewed under the categories of Psychological Modelling, Learning, Planning

Abstract

In planning a proof, a student searches through a space of inferences leading forward from the givens of the problem and backward from the to-be-proven statement. One dimension of growth of expertise is that students become more tuned in the search of this problem space. This can be shown to result from the application of various learning operators to production embodiments of the inference rules. Rules are evaluated after the solution of a problem according to whether they led to or led away from the solution. Rules that contributed to a solution are strengthened and an attempt is made to formulate general versions of these rules that will apply in other situations. Rules that led away from the solution are weakened and a discrimination process is evoked to try to add features to the rules that will try to restrict them to the correct circumstances of application. Composition is a learning process that collapses successful sequences of rule operations into single macro-rule productions. There is also a process that converts the backward reasoning rules formed by composition into forward reasoning rules. The effect of these learning processes is to put into production condition tests for problem features that are heuristically predictive of the rule's success.

*This paper draws heavily from my collaboration with James Greeno, Paul Kline, and David Neves. However, it should not be construed, despite the frequent use of "we" that they are in agreement with all the proposals put forth. The research described in this report was supported by contract N00014-78-C-0725 from the Office of Naval Research and is currently being supported by IST80-15357 from the National Science Foundation. I would like to thank Gary Bradshaw and Matt Lewis for their comments on the paper.

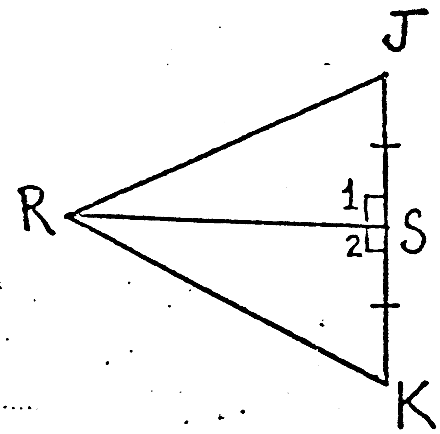
I have been involved in research (Anderson, Greeno, Kline, & Neves, 1981; Neves & Anderson, 1981) to characterize the organization of various proof skills possessed by high school students in geometry and to identify how these proof skills are acquired. In this paper I will concentrate on the skill involved in planning a proof to a geometry problem and, in particular, how the search for such a plan improves with practice. The direct goal in this research is to provide an accurate psychological model of how high school students learn to do geometry problems by doing geometry problems. A perhaps-not-incidental by-product is a set of ideas for how learning mechanisms might be used to guide problem-solving. All this research is done in the context of a general production system simulation of human cognition called ACT.

Organization of the Task and the Skill

The planning process we are trying to model is how students find a sequence of legal deductions that allow them to derive a to-be-proven statement from the givens. Figure 1 illustrates a triangle congruence problem which is sinple but nonetheless is challenging for the just beginning student. It is taken from the textbook we have been studying (Jurgensen, Donnelly, Maier, & Rising, 1975). Figure 2 illustrates the attempt of one of our subjects to solve this problem. First he tried to use the SSS method which worked on the previous problem. However, he noted that there seemed no way to get $\overline{RJ} \cong \overline{RK}$ and turned to side-angle-side. He immediately saw $\overline{JS} \cong \overline{KS}$ would provide one side and $\overline{RS} \cong \overline{RS}$ another side. He had a little difficulty seeing the included angle. His protocol at the critical point (after identifying the two congruent segments) reads "But where would $\angle 1$ and $\angle 2$ are right angles come in . . . Oh, I see how they work." This evidence, consistent with the rest of the protocol, shows that he did not see that right angles implied angle congruence until he needed angle congruence for the SAS postulate. At this point his plan was complete. He had some difficulty converting it into a legal two column proof (e.g., remembering that the reason that justified $\overline{RS} \cong \overline{RS}$ was called the "reflexive property of congruence") but there was no more planning in his protocol.

Figure 2 illustrates in simple form the backward search that is typical of novices in geometry and other domains (Larkin, McDermott, Simon, & Simon, 1980). Our simulation program plans in part by generating such a planning tree. In this tree there are disjunctions of methods to accomplish a goal (e.g., either SSS or SAS to prove triangle congruence) and each method can break down into a conjunction of subgoals (i.e., two sides and an included angle). Novices and our simulation (with a

Figure 1

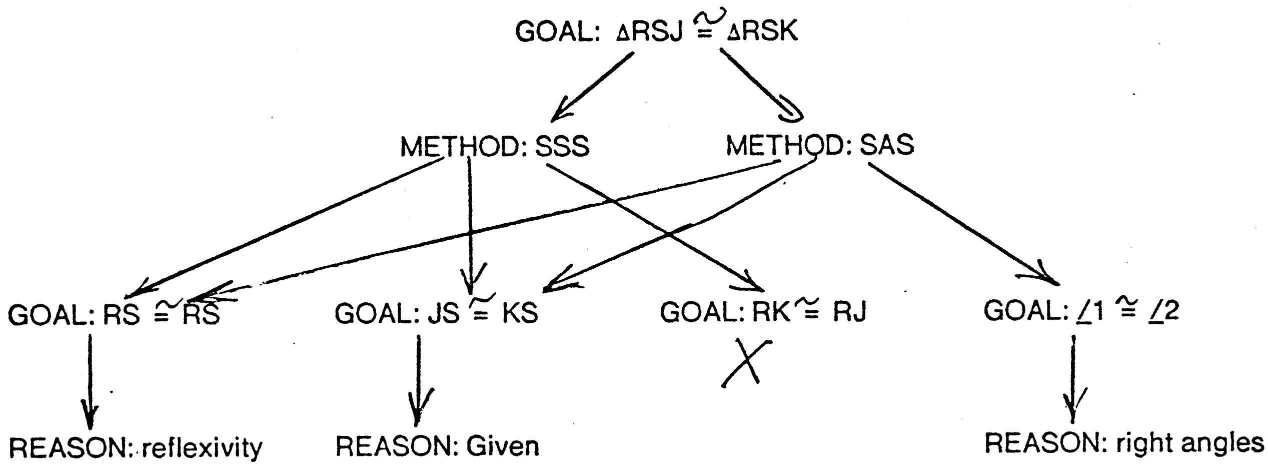


Given: $\angle 1$ and $\angle 2$ are right angles

$\overline{JS} \cong \overline{KS}$

Prove: $\triangle RSJ \cong \triangle RSK$

Figure 2


GOAL: ΔRSJ ≅ ΔRSK
METHOD: SSS
METHOD: SAS
GOAL: RS ≅ RS
GOAL: JS ≅ KS
GOAL: RK ≅ RJ
GOAL: ∠1 ≅ ∠2
REASON: reflexivity
REASON: Given
REASON: right angles

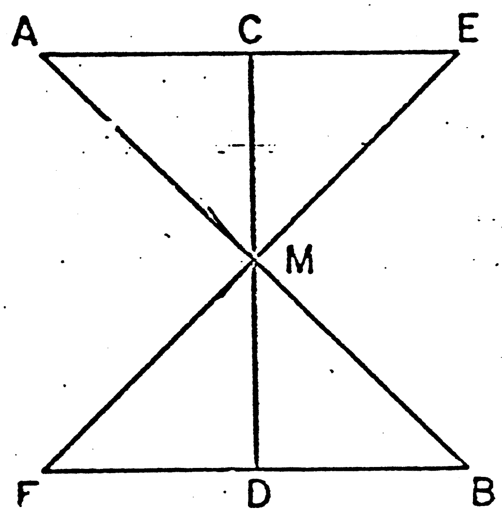
novice knowledge base) tend to search such a proof tree in a depth-first manner.

In more experienced students one sees forward inference from the givens. For instance, in problem 1 a student with some experience would likely recognize that $\angle 1$ and $\angle 2$ are congruent before he had consciously chosen the side-angle-side method. Potentially, geometry problems could be solved by pure forward search, but many potential forward inferences (e.g., those authorized by the reflexive rule) would be wasted. Optimal performance will arise from a mixture of forward and backward search. Figure 3 gives a problem that nicely illustrates the trade-off between forward and backward search. The majority of the subjects we have looked at in solving this problem (all at some intermediate level of skill) first reasoned forward to the inference the $\Delta AMC \cong \Delta BMD$ without knowing how they would use the fact. Then they worked backward to a proof plan that involved this forward inference. Our simulation at one setting did the same (see Anderson et al., 1981 for details).

It has been documented in other domains such as physics (Larkin, McDermott, Simon, & Simon, 1980) that the proportion of backward search decreases and the proportion of forward search increases with expertise. In our simulation the amount of forward inference depends on the existence of production rules that will make the forward inferences, on their strength, and whether various tests on their applicability are met. It is typical of our simulation that it will generate some set of forward inferences and then settle into a backward reasoning mode to complete the proof plan. This also seems typical of students who frequently start off marking some set of forward inferences on the diagram. Forward inferences tend to precede backward inferences in our simulation because they require less coordination and can therefore be more quickly executed.

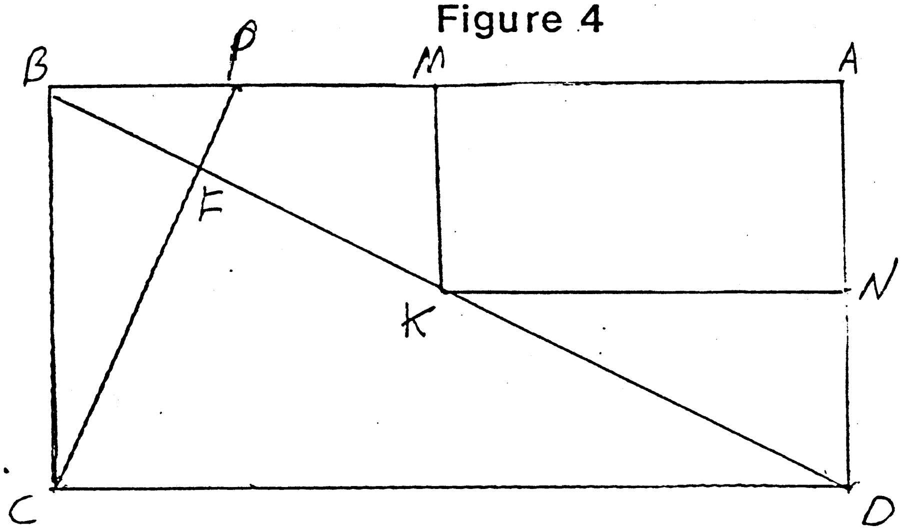
It is clear that either in forward or backward inference mode, there is a serious search problem for students. In forward inference mode one wants to only make those inferences that will play an essential role in the final proof. In backward inference mode one wants to pursue only those methods that lead to success. Neither our students nor the simulation are always successful in their search. However, it seems clear that one dimension of expertise is the ability to make more judicious decisions about the paths to search. The main focus of this paper is how that expertise is gained. (If the reader would like a problem likely to ~~pose search problems~~ create the experience of search for his level of expertise, I suggest he

Figure 3



GIVEN: M is the midpoint of $\overline{AB}$ and $\overline{CD}$
PROVE: M is the midpoint of $\overline{EF}$

Figure 4



Given

$m\angle BCD = 90^\circ$
$\overline{BA} \parallel \overline{CD}$
$\overline{BC} \parallel \overline{AD}$
$\overline{KN} \parallel \overline{CD}$
$\overline{KN} \parallel \overline{BA}$
$\overline{MK} \parallel \overline{DA}$

$\overline{BPMA}$ $\overline{BFKD}$ $\overline{CFP}$

$\overline{BP} \cong \overline{PM}$ $\overline{AM} \cong \overline{BC}$
$BM \cong \overline{MA}$

Prove: $m\angle BFC = 90^\circ$

consider solving the problem in Figure 4.)

The central theses of this paper are that there are certain features of a problem that are predictive of the success of a particular inference pathe and that the student learns the correlations (through proving problems). Some correlations between problem features and inference rules are logically determined. So, for instance, a student will learn that if he is trying to prove two triangles congruent and they both involve right angles, it is likely that he should try a right angle postulate. Other correlations between problem features and inference rules reflect more about biases in problem construction than any logical necessity. So, for instance, a student learns that if he sees a triangle that looks as if it is isosceles, it is likely that he will want to prove that it is isosceles. Whatever the reason for the correlation between features and inference methods, the student can use these feature-method correlations as heuristics to guide search. This paper is concerned with methods that can discover and exploit these correlations.

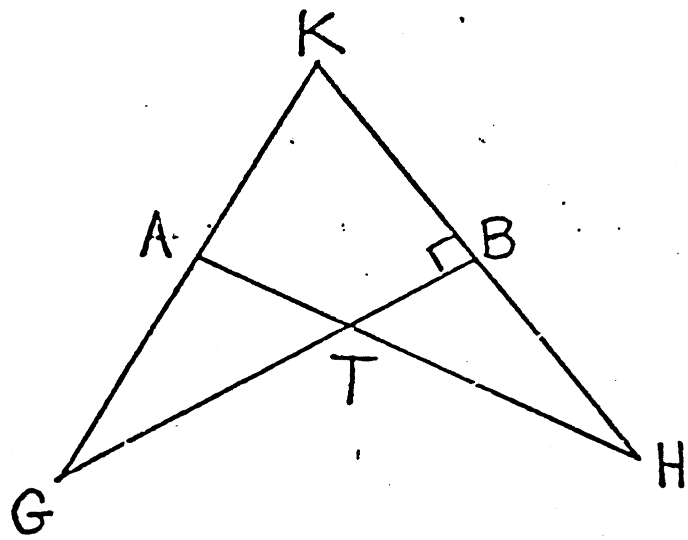
It is the character of heuristics that they should not always work and that it is possible to create problems that will violate these heuristics and which will, as a consequence, create difficulties. Figure 5 illustrates such a problem which occurred in the textbook we were using. The problem appears as an exercise immediately after the section that presents the hypotenuse-leg postulate for right-angle triangles. The majority of the subjects we have given this problem to report reasoning from the fact that $\angle H$ and $\angle K$ are complementary to the fact that $\angle A$ is a right angle. Then they can apply the hypotenuse-leg theorem. However, a simpler proof exists by simply noting the two triangles share $\angle K$ and then applying the side-angle-side proof. However, subjects are led by various heuristics such as (1) Problems tend to use the postulates introduced in the section; (2) If right-angles are mentioned and it is a triangle congruence problem, use a right-angle postulate; (3) Use all the givens in a problem. Students are generally not instructed as to such heuristics; they have picked them up by example.

Learning Mechanisms

I will discuss six methods for using the experience of past problems to improve search on current problems. We have worked on each method in our computer simulation and have reason for believing

Figure 5

GIVEN: ∠GBK is a right ∠
∠H is comp to ∠K
$\overline{AK} = \overline{BK}$
$\overline{GK} = \overline{HK}$
PROVE: ΔGBK = ΔHAK



that each is found in high school students. The first, analogy to prior problems, is somewhat distinct from the rest and will therefore be treated separately. The other five are principles concerned with extracting general and reliable rules from examples. They are the principles of rule evaluation, generalization, discrimination, composition, and forward inference formation. These last five make critical use of the production system architecture in which the simulation is implemented. The first, analogy, does not.

Analogy

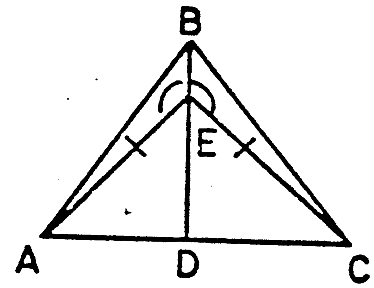
Despite the fact that its role is somewhat singular in our theory, our protocols are rich in evidence of successful problem-solving by analogy and many more attempts to use analogy. In the theory, analogy involves two processes. First, there is the noticing of the similarity between the specifications of a current problem and the specifications of a previous problem. Second, an attempt is made to map the solution to the previous similar problem to the current problem. The first process in our protocol is sufficiently rapid that it cannot be decomposed into substeps. A student will typically simply announce after reading the problem--"This is similar to Problem X." We have not been able to identify any instances where this Problem X occurred any earlier than in the previous day's lesson. So, there appears to be important memory limitations to the range of similarity noticing in analogy process.

We have implemented a partial graph matching process to model this similarity noticing. This partial matching process is also used in our work on generalization. The basic idea is an attempt to identify subgraphs on which the problems overlap. An early version of this is described in Anderson, Kline, and Beasley (1979, 1980) and a more advanced version by Kline (19xx). The ideas are variations on techniques suggested by Hayes-Roth and McDermott (1976) and Vere (1977).

Such a similarity detection mechanism is very much influenced by how the problems are represented. Consider Figure 6. In terms of many features such as shape and orientation, problems (a) and (c) are more similar than (a) and (b). However, it turns out that the more profitable similarities exist between (a) and (b). Many of the unsuccessful attempts to use analogy in our protocols can be accounted for by subjects being distracted by such superficial similarities.

Figure 6

(a)

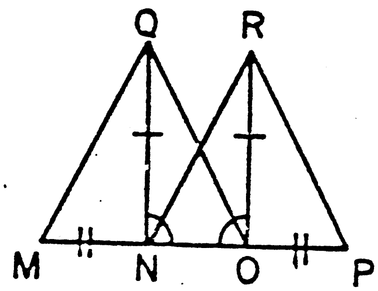


Given: $\overline{AE} \cong \overline{EC}$

$\angle BEA \cong \angle BEC$

Prove: $\triangle ABD \cong \triangle CBD$

(b)



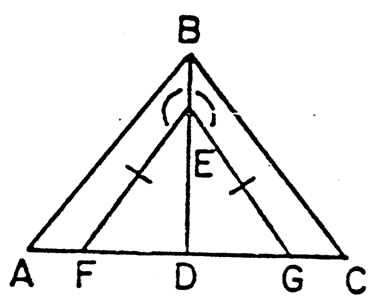
Given: $\overline{QN} \cong \overline{OR}$

$\angle QNO \cong \angle RON$

$\overline{MN} \cong \overline{OP}$

Prove: $\triangle MQO \cong \triangle PRN$

(c)



Given: $\overline{FE} \cong \overline{GE}$

$\angle BEF \cong \angle BEG$

$\overline{AB} \parallel \overline{FE}$

$\overline{BC} \parallel \overline{EG}$

Prove: $\triangle ABD \cong \triangle CBD$

In contrast to the rapid similarity-detection, the efforts to map a proof from one problem to another are quite long and definitely analyzable into substeps. It seems that the student has transformed his initial problem space into a new problem space of finding the mapping. We have not in our simulation work modelled this mapping process systematically. Figure 7 illustrates one of the more striking examples of failure of the mapping. The student noted the similarity between the two problems and proceeded to copy the proof to one problem over to the other. The first line for part (a) read RO = NY so analogously he wrote AB > CD for part (b). The second line for part (a) read ON = ON so analogously he wrote BC > BC for part (b)! His semantic sensibilities detected the problem; he abandoned the attempt to use the analogy; and proceeded to solve part (b) on its own.

While these two examples illustrate analogy by showing how it can fail, it is clear that it succeeds more often than not. One major problem with it is that it does not provide any permanent benefit as seen by the fact that all analogies are to problems encountered in the current or previous day. It may be that formulating analogies causes more permanent operators to be formed. The generalization process that will be described could apply after solution by analogy although solution by analogy is not a pre-requisite to generalization.

Rule Evaluation

The core of our simulation is a set of production rules for making forward and backward inferences. Below I illustrate, in informal syntax, productions embodying the SAS rule for forward and backward inference.

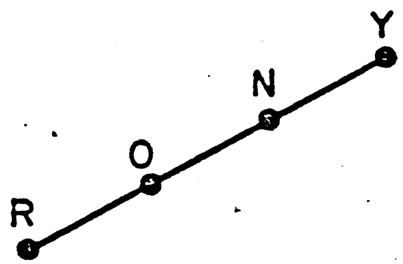
IF $\overline{XY} \cong \overline{UV}$, $\overline{YZ} \cong \overline{VW}$, and $\angle XYZ \cong \angle UVW$
THEN $\triangle XYZ \cong \triangle UVW$

IF the goal is to prove $\triangle XYZ \cong \triangle UVW$
THEN set as subgoals to prove $\overline{XY} \cong \overline{UV}$, $\overline{YZ} \cong \overline{VW}$, and $\angle XYZ \cong \angle UVW$

Other more elaborate production embodiments of these rules are also possible. The simulation keeps a record of the rules it applied in working on a problem. By comparing this record with the final proof plan it can determine which choices of proof method in working backwards were successful and which were mistakes. A little care is required here: Suppose a goal is set to prove two angles congruent by showing they are corresponding parts of congruent triangles. Suppose, all methods

Figure 7

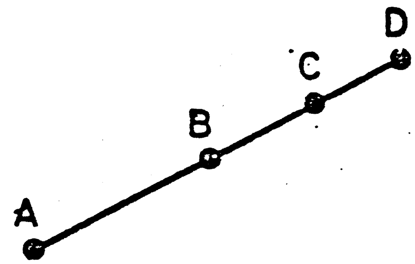
(a)



Given: $RO = NY$, $\overline{RONY}$
Prove: $RN = OY$

$$RO = NY$$
$$ON = ON$$
$$RO + ON = ON + NY$$
$$\overline{RONY}$$
$$RO + ON = RN$$
$$ON + NY = OY$$
$$RN = OY$$

(b)



Given: $AB > CD$, $\overline{ABCD}$
Prove: $AC > BD$

$$AB \geq CD$$
$$BC \geq BC$$

! ! !

tried for proving congruent triangles fail and the angle congruence is eventually proven by resorting to supplementary angle postulate. The mistake is not in the methods attempted for proving the triangles congruent rather the mistake was in setting the subgoal of triangle congruence. Forward inferences can be classified as successful if they figure in the final proof and as mistaken otherwise.

Success and error classifications are used by the learning mechanisms to be described shortly, but it is also used to simply strengthen or weaken the rules responsible for the decisions. The mechanisms for strengthening and weakening a production and the impact of production strength on conflict resolution has been described elsewhere (Anderson, Kline, & Beasley, 1979). However, it is important to note that disastrous results will not occur if a bad rule is formulated since the strength evaluation mechanism will separate out successful from unsuccessful rules and eventually only the former will be selected in conflict resolution. This means that we do not have to be concerned that the learning mechanisms always be correct in the production rules they formulate.

Generalization

Generalization attempts to extract common features of two instances and successfully apply the same inference method. This is done by testing for similarity between the problem descriptions before the rule of inference applies. Consider problems (a) and (b) of Figure 6. In both cases, the initial step involves setting as a subgoal to prove congruent triangles that overlap with the to-be-proven-congruent triangles. The representation of the state of knowledge for problem (a) at the point of setting this subgoal might involve the following clauses:

1. The goal is to prove $\triangle ABD \cong \triangle CBD$
2. $\triangle ABD$ contains $\triangle AEB$
3. $\triangle CBD$ contains $\triangle CEB$
4. $\overline{AE} \cong \overline{EC}$
5. $\angle BEA \cong \angle CEA$
6. B is at top
7. $\triangle ABC$ contains $\triangle ABD$
8. $\triangle ABC$ contains $\triangle CBD$

Similarly, the state of knowledge for problem (b) when this subgoal is set might be:

1. The goal is to prove $\triangle MQO \cong \triangle PRN$
2. $\triangle MQO$ contains $\triangle NQO$
3. $\triangle PRN$ contains $\triangle ORN$
4. $\overline{NQ} \cong \overline{OR}$
5. $\angle QNO \cong \angle RON$
6. $\overline{MP}$ is a horizontal line at the base
7. Q is at the top
8. R is at the top

These two states of knowledge can be generalized by extracting what they have in common. This generalized situation can be assigned the common action of subgoal proof of the contained triangles by the production:

IF the goal is to prove $\triangle XYZ \cong \triangle UVW$
and $\triangle XYZ$ contains $\triangle SYZ$
and $\triangle UVW$ contains $\triangle TVW$
and $\overline{SY} \cong \overline{TV}$
and $\angle YSZ \cong \angle VTW$
THEN set as a subgoal to prove $\triangle SYZ \cong \triangle TVW$

The extraction of such similarities is described in Anderson, Kline, and Beasley (1980) and Anderson and Kline (1979) and is similar to ideas proposed earlier by Hayes-Roth and McDermott (1976) and by Vere (1977). As noted earlier, generalization involves the same mechanisms used in similarity detection for analogy. The above example illustrates how it might be used to extract from examples the principle of chaining the goal of proving triangle congruence to a subgoal of proving the triangle congruence of overlapping triangles. Note that the generalization preserves features specific to the two examples that are predictive of the method's success--namely, that parts of the overlapping triangles are congruent.

The evidence is quite clear that subjects do extract from examples methods that work over a class of examples containing features. The overlapping triangles rule above is one although it usually appears to be more general in that problem solvers will try to chain to overlapping triangles whenever they contain one or two congruent pieces--not a specific side and angle. A more general production such as this could derive from the one above by further generalization (with appropriate representational assumptions).

Although students do have these general rules without a doubt; it is unclear that they emerge by the generalization mechanissm suggested above. As an alternative, they might derive by a retrospective analysis of a single problem rather than a generalization between two. Our protocol data cannot inform us on this issue and we are tooling up to do the right kinds of controlled experiments. Work on extraction of object categories (Anderson & Kline, 1979; Elio & Anderson, 1981) has provided good evidence for a generalization process in that domain.

Discrimination

The initial rules that a system has comes from more-or-less direct encodings of postulates. So, for instance, the SSS postulate can lead to a rule of the form

IF the goal is to prove $\Delta XYZ \cong \Delta UVW$
THEN set as subgoals to prove $\overline{XY} \cong \overline{UV}$, $\overline{YZ} \cong \overline{VW}$, and $\overline{ZX} \cong \overline{WU}$

The problem with such productions is that their conditions are too general and do not lead to selectivity of search. It is also the case that the generalization process itself might produce overly general productions. Overly general rules can be restricted with by a discrimination mechanism which compares successful and unsuccessful applications of a production, tries to determine the features which distinguish the successful applications, and proposes new productions derived from the old but which contain these distinguishing features in their conditions. Again, the details of the discrimination procedure have been described in Anderson and Kline (1979), Anderson, Kline, and Beasley (1980) and I will simply describe here their application to the geometry domain.

Consider the representation in Figure 2 of the student's planning for the problem in Figure 1. After completing this problem the learning program would identify the attempt to use SSS as a mistake and SAS as the correct method. Comparing this situation to the previous problem that was successfully solved by SSS, the program would note that this problem differs only in the fact that right angles are mentioned. Thus, it could propose the following discrimination:

P1: IF the goal is to prove $\triangle XYZ \cong \triangle UVW$
and $\triangle XYZ$ contains no right angles
and $\triangle UVW$ contains no right angles
THEN set as subgoals to prove $\overline{XY} \cong \overline{UV}$, $\overline{YZ} \cong \overline{VW}$, and $\overline{ZX} \cong \overline{WU}$

The discriminating clauses are found by locating additional clauses in the data base that constrain the variables. The above discrimination is probably too specific and should not be limited to right angles. Through generalization with other productions that do not involve right angles, the no right angle requirement could be replaced by the requirement that no angles be mentioned.

It is also possible to form a discrimination of a different variety by embellishing the SAS rule to encode what is distinctive about the current situation. This can lead to a production of the following sort:

P2: IF the goal is to prove $\triangle XYZ \cong \triangle UVW$
and $\triangle XYZ$ contains a right angle triangle $\angle XYZ$
and $\triangle UVW$ contains a right angle triangle $\angle UVW$
THEN set as goals to prove $\overline{XY} \cong \overline{UV}$, $\overline{YZ} \cong \overline{VW}$, and $\angle XYZ \cong \angle UVW$

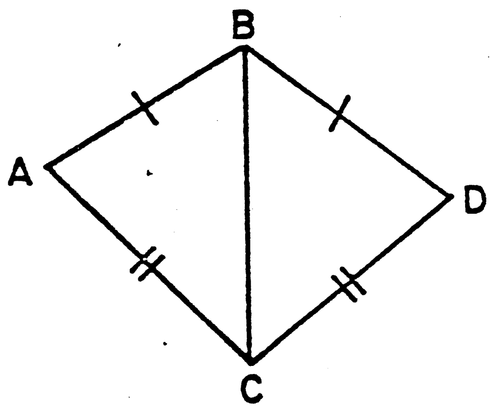
This type of discrimination was not produced in our original ACT simulation but has proven useful in some of our more recent, special purpose simulations (Anderson, 1981).

As in the case of generalization, the fact is indisputable that subjects form discriminations on their original rules. Indeed, one subject articulated a rule essentially identical to P1 after the history illustrated in Figure 2. However, again as in the case of generalization, what is unclear is whether these discriminations are achieved by the mechanisms described here. Again, that issue awaits more detailed experimental research.

Composition

Neves and Anderson (1981), developing ideas put forth by Lewis (1978), applied a learning mechanism called composition to proof generation in geometry. The basic idea behind the composition mechanism is to package sequences of production steps into single operators. A somewhat similar idea in the domain of logic proofs has been advanced by Smith (198x). Figure 8 illustrates one of the problems where we applied this mechanism. The first pass of this system over the problem was accomplished by a sequence of five productions. The first production to apply in

Figure 8



Given: $\overline{AB} \cong \overline{DB}$
$\overline{CA} \cong \overline{CD}$
Prove: $\angle A \cong \angle D$

solving this problem was:

P1: IF the goal is to prove $\angle X \cong \angle U$
and $\angle X$ is part of ΔXYZ
and $\angle U$ is part of ΔUVW
THEN the subgoal is to prove $\Delta XYZ \cong \Delta UVW$

This production would set as a subgoal to prove $\Delta ABC \cong \Delta DBC$. At this point the following production applied:

P2: IF the goal is to prove $\Delta XYZ \cong \Delta UVW$
and $\overline{XY} \cong \overline{UV}$
and $\overline{ZX} \cong \overline{WU}$
THEN the subgoal is to prove $\overline{YZ} \cong \overline{VW}$

This production, applied to the situation in Figure 8 set as the subgoal to prove $\overline{BC} \cong \overline{BC}$ as a step on the way to using SSS. At this point the following production applied:

P3: IF the goal is to prove $\overline{XY} \cong \overline{XY}$
THEN this may be concluded by reflexivity

This production added $\overline{BC} \cong \overline{BC}$ and allowed the following production to apply:

P4: IF the goal is to prove $\Delta XYZ \cong \Delta UVW$
and $\overline{XY} \cong \overline{UV}$
and $\overline{YZ} \cong \overline{VW}$
and $\overline{ZX} \cong \overline{WU}$
THEN the goal may be concluded by SSS

where XY = AB, UV = DB, YZ = BC, VW = BC, ZX = CA, and WU = CD. This adds the information that $\Delta ABC \cong \Delta DBC$. Finally, the following production applied which recognizes that the to-be-proven conclusion is now established:

P5: IF the goal is to prove $\angle X \cong \angle U$
and $\Delta XYZ \cong \Delta UVW$
THEN the goal may be concuded because of congruent parts of congruent triangles

The composition process, operating on this sequence, produced a single production that served as a macro-operator:

P6: IF the goal is to prove $\angle A \cong \angle D$
and $\angle A$ is part of ΔABC
and $\angle D$ is part of ΔDBC
and $\overline{AB} \cong \overline{DB}$
and $\overline{CA} \cong \overline{DC}$

THEN conclude $\overline{AB} \cong \overline{AB}$ by reflexivity
and conclude $\Delta ABC \cong \Delta DBC$ by SSS
and conclude the goal because of congruent parts of congruent triangles

The variables in this production have been named to correspond to the terms in Figure 8 for purposes of readability. This production would immediately recognize the solution to a problem like that in Figure 8. This composition is achieved basically by adding together the conditions of the five original productions and making them the condition of the composed production; adding together the actions and making these the action of the composed production, editing out the unnecessary or redundant clauses in the composed production.

The details for the editing of the unnecessary clauses are given in Neves and Anderson (1981) but basically they involve (a) eliminating clauses from the conditions of productions late in the sequence which are satisfied by the actions of productions early in the sequence and (b) eliminating the setting and testing of goals which are met in the sequence. With respect to (b) composition serves the effect of collapsing several levels in the proof plan into a single level.

There is a question of when to evoke composition. The original implementation by Neves and Anderson was to use composition on any immediately contiguous sequence of productions. However, I think it more reasonable to have it apply to a sequence of productions related by manipulation of the same goals--as in the case just illustrated. These two definitions of production sequence need not yield the same sequences in the ACT system. It is quite possible for immediately contiguous productions not to share similar goals.

Creation of Forward Inferences

It is a feature of the composition production P6 that it summarizes what had been a multi-level goal tree. The system had started with the goal of proving two angles congruent, set a subgoal of proving two triangles congruent, set a subgoal of proving two sides congruent, and then proceeded to pop the goals finally achieving the highest level goal. It would be useful to have this available as a forward inference rule so that, should the situation appear again, the inferences can be made to embellish the problem representation. This can be achieved by dropping the goal specification from P6 (a similar idea was proposed by Larkin, 1981). The resulting production would be:

P7: IF $\angle A$ is part $\triangle ABC$
and $\angle D$ is part of $\triangle DBC$
and $\overline{AB} \cong \overline{DB}$
and $\overline{CA} \cong \overline{DC}$
THEN conclude $\overline{AB} \cong \overline{AB}$ by reflexivity
and conclude $\triangle ABC \cong \triangle DBC$ by SSS
and conclude $\angle A \cong \angle D$ because of congruent parts of congruent triangles

It is interesting to note that this production makes a reflexive inference in forward mode. To have a pure reflexive rule as a forward inference:

IF $\overline{AB}$
THEN $\overline{AB} \cong \overline{AB}$ by reflexivity

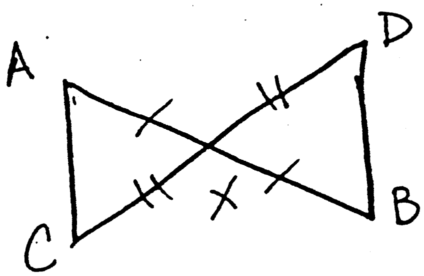
would be a sheer disaster since it would complicate the problem representation with many useless inferences. However, cast as part of a larger operator as above it is a very profitable forward inference.

To review, forward inferences can be made when composition creates a macro-operator which achieves a stated goal by a sequence of inferences that previously had involved the embedding of subgoals. The forward inference can be created from the composition by deleting the goal clause. It is useful to understand why one would only want to drop goal clauses from the macro-operators rather than the original working-backwards productions. The original productions are so little constrained that the goal clauses provide important additional tests of applicability. After a macro-operator is composed there are enough tests in the non-goal aspects of its condition to make it quite likely that the inferences will be useful. That is, it is unlikely to be an accident that the conjunction of tests are satisfied.

We do not have any evidence for P6 or P7 as specific inference rules--probably because the pattern in Figure 8 occurs but once in the textbook exercises. In contrast the pattern in Figure 9 or slight variants of it occur quite frequently in the chapter. The problem in Figure 9 can be solved by the following three productions.

P8: IF the goal is to prove $\triangle XYZ \cong \triangle UVW$
and $\overline{XY} \cong \overline{UV}$ and $\overline{YZ} \cong \overline{VW}$

Figure 9



GIVEN: $\overline{AX} = \overline{XB}$
$\overline{CX} = \overline{XD}$
$\overline{AXB}, \overline{CXD}$
PROVE: $\Delta AXC = \Delta BXD$

THEN set as a subgoal to prove $\angle XYZ \cong \angle UVW$

P9: IF the goal is to prove $\angle XYZ \cong \angle UYW$
and $\overline{XYW}$ and $\overline{UYZ}$
THEN this can be concluded by vertical angles

P10: IF the goal is to prove $\triangle XYZ \cong \triangle UVW$
and $\overline{XY} \cong \overline{UV}$, $\overline{YZ} \cong \overline{VW}$, and $\angle XYZ \cong \angle UVW$
THEN this can be concluded by SAS

Composing these three productions together we get:

P11: IF the goal is to prove $\triangle XYZ \cong \triangle UYW$
and $\overline{XY} \cong \overline{UV}$ and $\overline{YZ} \cong \overline{YW}$
and $\overline{XYW}$ and $\overline{UYZ}$
THEN conclude $\angle XYZ \cong \angle UYW$ by vertical angles
and conclude $\triangle XYZ \cong \triangle UYW$ by SAS

Deleting the goal clause we get the following forward inference:

P12: IF there are $\triangle XYZ$ and $\triangle UYW$
$\overline{XY} \cong \overline{UY}$ and $\overline{YZ} \cong \overline{YW}$
and $\overline{XYW}$ and $\overline{UYZ}$
THEN conclude $\angle XYZ \cong \angle UYW$ by vertical angles
and $\triangle XYZ \cong \triangle UYW$ by SAS

There is clear evidence for such a forward inference rule in some more advanced students. For them, the pattern in Figure 9 is something that will trigger the set of inferences even when it appears embedded in a larger problem. ~~What~~ However, we have poor evidence on what the exact origins are of this forward inference rule.

Final Points

We have described a set of mechanisms that might plausibly account for the growth in expertise of proof search. The mechanisms of generalization, discrimination, and composition were shown to produce more tuned judgments in backward search--at least in specific cases. These mechanisms could also be used to produce more tuned and powerful forward inference rules. The final learning method involved dropping goal clauses of composed backward productions to create forward reasoning productions. This serves to produce the frequent observation of a shift in search from backward to forward inference with growth of expertise. There are certainly other aspects to learning in problem solving besides this tuning of search, but this is an important aspect. The apparent ease with which this tuning can be produced in a production system architecture is evidence for that architecture.

There are two major issues that need to be pursued. One, as noted throughout the paper, is the detailed empirical verification of these mechanisms. The second is a sufficiency proof of their operation by simulating a student's course through a textbook. While we have worked up individual examples of successful tuning by these learning mechanisms, we have not done the large scale simulation to show that their cumulative effect after hundreds of problems will match the degree of tuning we see in the typical student. We intend to pursue this and I am reasonably optimistic given that we have achieved success with such large-scale simulations of our learning mechanisms in the domain of concept formation (Anderson & Kline, 1979) and syntax acquisition (Anderson, 1981).

References

Anderson, J.R. A theory of language acquisition based on general learning principles. IJCAI, 1981, submitted.

Anderson, J.R., Greeno, J.G., Kline, P.J., & Neves, D.M. Acquisition of problem-solving skill. To appear in J.R. Anderson (Ed.), Cognitive Skills and their Acquisition. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Anderson, J.R. & Kline, P.J. A learning system and its psychological implications. In the Proceedings of the Sixth International Joint Conference on Artificial Intelligence, 1979.

Anderson, J.R., Kline, P.J., & Beasley, C.M. A general learning theory and its application to schema abstraction. In G.H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 13. New York: Academic Press, 1980.

Anderson, J.R., Kline, P.J., & Beasley, C.M. Complex learning processes. In R.E. Snow, P.A. Federico, & W.E. Montague (Eds.), Aptitude, Learning, and Instruction: Cognitive Process Analyses. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Elio, R. & Anderson, J.R. Effects of category generalizations and instance similarity on schema abstraction. Journal of Experimental Psychology: Human Learning and Memory, in revision.

Hayes-Roth, F. & McDermott, J. Learning structured patterns from examples. Proceedings of the Third International Joint Conference on Pattern Recognition, 1976, 419-423.

Jurgensen, R.C., Donnelly, A.J., Maier, J.E., & Rising, G.R. Geometry. Boston, MA: Houghton Mifflin Co., 1975.

Larkin, J. Enriching formal knowledge: A model for learning to solve textbook physics problems. In J.R. Anderson (Ed.), Cognitive Skills and their Acquisition. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. Models of competence in solving physics problems. Cognitive Science, 1980, 317-345.

Lewis, C.H. Production system models of practice effects. Unpublished Ph.D. dissertation, University of Michigan, 1978.

Neves, D.M. & Anderson, J.R. Knowledge compilation: Mechanisms for the automatization of cognitive skills. To appear in J.R. Anderson (Ed.), Cognitive Skills and their Acquisition. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Smith, R.L. Modelling student acquisition of problem-solving skills: The student's interpretation of the teaching environment.

Vere, S.A. Induction of relational productions in the presence of background information. Proceedings of the Fifth International Joint Conference on Artificial Intelligence. Boston, MA, 1977, 349-355.