COGNITIVE PSYCHOLOGY 17, 248–294 (1985)

Why Are Some Problems Hard? Evidence from Tower of Hanoi

K. KOTOVSKY

Community College of Allegheny County

AND

J. R. HAYES AND H. A. SIMON

Carnegie–Mellon University

This paper analyzes the causes for large differences in difficulty of various isomorphic versions of the Tower of Hanoi problem. Some forms of the problem take 16 times as long to solve, on average, as other versions. Since isomorphism rules out size of task domain as a determinant of relative difficulty, these experiments seek and find causes for the differences in features of the problem representation. Analysis of verbal protocols and the temporal patterning of moves allows the problem-solving behavior to be divided into exploratory and final-path phases. Entry into the final-path phase depends on acquisition of the ability to plan pairs of moves, an achievement made difficult by the working memory load it entails. This memory load can be reduced by automating the rules governing moves, either through problem exploration or training. Once automation has occurred, the solution is obtained very rapidly. Memory load is also proposed as the locus of other differences in difficulty found between various problem representations.

INTRODUCTION

A major turning point in the study of problem solving was the development of the information-processing approach, which analyzed problems by means of a "problem space" representing the various states of knowledge of a problem solver along the way toward a solution, the transformations between allowable states, and the operators that accomplished those transformations (Newell & Simon, 1972). From the problem-space approach there emerged the idea that size of problem space (number of branches at each node and depth of search to a solution node) was a principal determinant of problem difficulty. This is certainly true (and mathematically demonstrable) if problems are solved by random

This research was supported by National Science Foundation Grant BNS-78-25033. Requests for reprints should be sent to K. Kotovsky, Behavioral Sciences Department, Community College of Allegheny County, 808 Ridge Avenue, Pittsburgh, PA 15212.

0010-0285/85 $7.50

trial-and-error search. However, it does not explain why some problems with quite small problem spaces are difficult for intelligent people. The Missionaries and Cannibals (Hobbits and Orcs) problem has a problem space of only 16 nodes, and monster problem versions of the three-disk Tower of Hanoi problem, only 27 nodes. Both problems are known to be difficult for human subjects who encounter them for the first time.

The Tower of Hanoi problem has long been a major task environment for work in problem solving. This research has included:

- A formal analysis of the problem and its problem space (Nilsson, 1971).
- An analysis of the effect of various task instructions on problem difficulty (Gagné & Smith, 1962).
- Application of a general problem-solving model (GPS) to the problem (Newell & Simon, 1972).
- Analysis of the difficulty engendered by the depth of the goal stack for four-, five-, and six-disk versions of the problem (Egan & Greeno, 1974).
- Generation and formal analysis of different solution strategies in the form of production systems (Simon, 1975).
- A production system model that accounts for the latencies found in four-, five-, and six-disk problems (Karat, 1982).
- The detailed analysis, using production systems, of one subject's evolution of progressively more useful and efficient solution strategies as she worked on the problem (Anzai & Simon, 1979).
- A demonstration of the physiological distinction between procedural and declarative knowledge in the ability of amnesiac patients to learn how to perform a five-disk version of the problem flawlessly while evincing no knowledge of having ever seen or solved the problem before (Cohen & Corkin, 1981).

In 1974, Hayes and Simon published the first of a series of papers on subjects' performances solving various isomorphs of the Tower of Hanoi problem. These isomorphs used different cover stories but always involved a problem space identical in size, branchiness, and minimum solution path length to a three-disk Tower of Hanoi problem. In addition, the rules or operators that accomplished transformations between allowed problem states were identical in number, relevance, and restrictiveness to the rules in the Tower of Hanoi problems.

Despite the structural identity of these isomorphs, subjects solving various versions exhibited large and systematic differences in the relative difficulty they experienced with two broad classes of isomorphs. These two classes have been labeled ***Move*** (or *Transfer*) problems and ***Change*** problems, respectively (Hayes & Simon, 1974, 1977; Simon & Hayes, 1976), on the basis of whether successive problem transformations require moving an object from one spatial location to another, or changing some property of an object that remains at a fixed location. The first example in Fig. 1 is a Move problem, and the second is a Change problem, both using a Monster cover story.

As can be seen in that figure, the Move problem, TA, involves three globes of different sizes that are passed back and forth between three monsters, which are also of varied sizes. The rules determining move legality are based on the relative sizes of the three globes. The problem corresponds to a five-step Tower of Hanoi problem with the three disks (globes) distributed over the three pegs (monsters) at both start and finish. The Change problem, CA′, is also a five-step problem involving three globes of different sizes that can be changed from one size to another. Each globe is held by a monster that is also one of three sizes. The rules in the Change problem depend on the relative size of the monsters, rather than the globe sizes. Here the monsters correspond to the disks of the Tower problem, and the globes correspond to the pegs. Hence, the problems are structurally isomorphic to each other and to the corresponding three-disk Tower of Hanoi problem.

Both classes of Monster problems are very much more difficult than the original Tower of Hanoi problem (a *Move* problem) with its physical embodiment of the solution rules. In addition, the Change problems, in a variety of forms, are generally twice as difficult as otherwise comparable Move problems (Hayes & Simon, 1977). These findings are summarized in Table 1. Using the Tower of Hanoi three-disk problem as a standard, it takes 3 times as long to solve an Acrobat problem, 5 times as long to solve a Reverse-Acrobat problem, 8 times as long to solve a Monster Move problem, and 16 times as long to solve a Monster Change problem.[1]

These results, along with other findings, raise two related questions:

- What is the source of the large (2:1) differences in difficulty between Monster Change and Move isomorphs, and the even larger (16 and 8 to 1, respectively) differences between these and the isomorphic Tower of Hanoi problem?
- What properties of the problems (or more particularly, the problem representations) affect transfer of training between isomorphs?

Previous research on Tower of Hanoi isomorphs (see, for example,

[1] The Acrobat and Reverse-Acrobat problems are Move problems that describe acrobats (disks) jumping onto each others' shoulders as they stand on three flagpoles (pegs). A larger acrobat may not land on a smaller in the Acrobat problem, and a smaller may not land on a larger in the Reverse-Acrobat problem. These isomorphs are described more fully in Experiment 3.

A Monster Move Problem - TA

Three five-handed extra-terrestrial monsters were holding three crystal globes. Because of the quantum-mechanical peculiarities of their neighborhood, both monsters and globes come in exactly three sizes with no others permitted: small, medium, and large. The small monster was holding the large globe; the medium-sized monster was holding the small globe; and the large monster was holding the medium-sized globe. Since this situation offended their keenly developed sense of symmetry, they proceeded to transfer globes from one monster to another so that each monster would have a globe proportionate to its own size.

Monster etiquette complicated the solution of the problem since it requires that:

1. Only one globe may be transferred at a time;

2. If a monster is holding two globes, only the larger of the two may be transferred; and,

3. A globe may not be transferred to a monster who is holding a larger globe.

By what sequence of transfers could the monsters have solved this problem?

[Your first goal should be to take care of the small monster (i.e., to get him the right sized globe).] This appeared in the *hint* condition only.

A Monster Change Problem - CA'

Three five-handed extra-terrestrial monsters were holding three crystal globes. Because of the quantum-mechanical peculiarities of their neighborhood, both monsters and globes come in exactly three sizes with no others permitted: small, medium, and large. The small monster was holding the medium-sized globe; the medium-sized monster was holding the large globe; and the large monster was holding the small globe. Since this situation offended their keenly developed sense of symmetry, they proceeded to shrink and expand the globes so that each monster would have a globe proportionate to its own size.

Monster etiquette complicated the solution of the problem since it requires that:

1. Only one globe may be changed at a time;

2. If two globes have the same size, only the globe held by the larger monster may be changed; and

3. A globe may not be changed to the same size as the globe of a larger monster.

By what sequence of changes could the monsters have solved this problem?

[Your first goal should be to take care of the small monster (i.e., to get his globe to the right size).] This appeared in the *hint* condition only.

FIG. 1. Problem isomorphs TA and CA′ for both the standard and the hint conditions.

Hayes & Simon, 1977) has focused attention on problem representations. Differences between problem isomorphs cannot be attributed to differences in problem structure, but rather must be attributed to differences in the manner in which the subject *images* or *models* or *thinks about* the problem. This shift of focus away from task domain structure and toward representation, together with recent work on representation in physics

TABLE 1
Difficulty of Various Tower of Hanoi Problem Isomorphs

Problem type	Average solution time (Minutes)	Source
Tower of Hanoi (Move problem)	1.83	Current Experiment 5B
Acrobat (Move problem)	5.63	Current Experiment 3A
Reverse Acrobat (Move problem)	9.51	Current Experiment 3B
Monster (Move problem)	13.95	Hayes & Simon (1977)
Monster (Change problem)	29.39	Hayes & Simon (1977)

problem solving (see, for example, Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980) and work on the effects of real world knowledge on isomorphic logic problems (Wason & Johnson-Laird, 1972), has defined a new set of research issues.

Two major components of the representation of a problem are (a) its problem space and (b) the move operators that allow for movement between states (Newell & Simon, 1972). Both components contain plausible potential determinants of problem difficulty. The first two hypotheses about problem difficulty we shall investigate relate to differences between isomorphs in the acquisition and utilization of the move operators and their restrictions (the problem *rules*); the next three explore representational issues.

1. *Differences in the ease of learning the rules for making moves.* While the number of rules and their formal structure is the same for all the problem categories, they do differ linguistically in ways that might affect the ease with which they can be learned. We will label this the **Rule Learning Hypothesis** and, as a corollary, will view similarity of sets of rules as a possible predictor of transferability of training: the **Rule Similarity/Transfer Hypothesis.** (It was identified previously as a likely explanation of transfer effects by Hayes and Simon, 1977.)

2. *Differences in the ease of applying the rules; particularly in conducting the tests that must be performed to determine legality of moves.* This will be labeled the **Rule Application Hypothesis.** This hypothesis was examined in Hayes and Simon (1977), where it was suggested as the source of the differences in problem difficulty between Move and Change problems.

3. *Differences in the consistency of real world knowledge with the problem rules.* Ease of rule assimilation and/or application might depend on the extent to which the rules are consistent with knowledge of the real world. This is designated the **Real World Knowledge Hypothesis.** This effect might be similar to the demonstrated superiority of the more familiar problem situations in Wason's selection task, and also in more

recent work on the effect of *real world* knowledge on the ease of understanding syntactically similar sentences (Wason & Reich, 1979; Wason & Shapiro, 1971).

4. *Differences in the memory loads imposed by various isomorphs.* Another factor that might affect both rule learning and application is the load that problems impose on working memory. Differences in memory load are likely to be a source of differences between the Tower of Hanoi problem (where to some extent the rules and rule restrictions are physically observable) and the Monster problem isomorphs where there is no such obvious external memory. These differences might be similar to the differences in difficulty between doing mental arithmetic and using paper and pencil. The hypothesis that external memory reduces difficulty by reducing memory load will be called the **External Memory Hypothesis.**

5. *Differences in the internal problem space representation elicited by the various cover stories.* Move problems might engender more spatially organized or more easily imaged problem representations and thus impose less of a memory load than do Change problems. This **Spatial Memory Load Hypothesis** must account for differences in difficulty even when subjects use external memory aids. One possibility is that subjects use such external representations for actually making moves, but have to rely on their internal representations in planning moves. General support for this view is offered by the results of an experiment in which subjects solved either a spatial or a temporal isomorph of a job-scheduling/office-layout problem. The subjects solving the spatial isomorph obtained higher scores in a shorter period of time than did the others (Carroll, Thomas, & Malhotra, 1980).

Our strategy is to analyze the problem situation and behavior in terms of the variables just listed, to generate experiments based on each type of influence, and thus to attempt to account for the large differences in problem difficulty and delineate some of the conditions for effective transfer of training from one problem to another.

EXPERIMENT 1: RULE LEARNING

The **Rule Learning Hypothesis** proposes differences in learning the problem rules as a source of the differences in difficulty between classes of problems. Experiment 1 was performed to test this hypothesis.

Method

Subjects. The subjects were students at a local community college who were paid for their participation in the experiment.

Procedure. The subjects were run individually. They were told by the experimenter that they would be asked to "learn some rules" that were typed on a sheet of paper, with one set of three rules on a sheet (Fig. 2). The subjects were instructed to study the rules aloud

CA Rules

This set of rules involves monsters changing the sizes of globes they are holding.

1. Only one globe may be changed at a time.

2. If two globes have the same size, only the globe held by the larger monster may be changed.

3. A globe may not be changed to the same size as the globe of a larger monster.

CA-2 Rules

This set of rules involved monsters changing the size of globes they are holding.

1. Only one globe may be changed at a time.

2. If two globes have the same size, only the darker globe may be changed.

3. A globe may not be changed to the same size as a darker globe.

TA Rules

This set of rules involves monsters transferring globes to and from other monsters.

1. Only one globe may be transferred at a time.

2. If a monster is holding two globes, only the larger of the two may be transferred.

3. A globe may not be transferred to a monster who is holding a larger globe.

TA-2 Rules

This set of rules involves monsters transferring globes to and from other monsters.

1. Only one globe may be transferred at a time.

2. If a monster is holding two globes, only the darker globe may be transferred.

3. A globe may not be transferred to a monster who is holding a darker globe.

FIG. 2. Problem rules for four isomorphs. CA and CA-2 are Change problems, TA and TA-2, Move problems.

until they felt they knew them, at which point they were to put the sheet face down and recite the rules to the experimenter.

The subjects were informed that they would be timed in learning the rules. The rules were then placed face down in front of the subjects, who were told to begin; simultaneously, a clock was started. When the subjects wished to recite the rules, they so indicated, put the sheet down, then turned to the experimenter and recited. The learning criterion was being able to recite the rules or a meaning-preserving paraphrase twice in succession. The dependent measure was time to learn the rules to criterion. A subject who forgot or distorted one of the rules was instructed to keep trying, at which point she/he would begin to restudy the rule sheet. The subjects were not told that these were rules for a problem.

Ten subjects were each given four sets of rules to learn.[2] The order of presentation was counterbalanced, although not perfectly. One subject experienced great difficulty with the task (taking 15 to 17 min to learn a rule set), and as a result, quit the experiment without finishing, and has been excluded from the analysis. (That subject's results on the two rule sets she completed are in agreement with the rest of the results.)

Results

The results for the remaining nine subjects are presented in Table 2. Median times for the subjects with Change rules were more than twice as long as medians for subjects with Move rules. An analysis of variance of the log transformed times showed the Move–Change difference to be significant [$F(1,8) = 5.6, p < .05$]. The effect of the color variable (moves or changes based on relative darkness), was not significant [$F(1,8) = .32$, $p = .59$], nor was the interaction between the two variables [$F(1,8) = 1.14$, $p = .32$]. These results are consistent with the major differences found repeatedly in past studies between Move and Change problems. The lack of significant interaction is not fully consistent, however, with the results of a previous experiment that found the Move problem TA-2 (which based move legality on relative darkness rather than relative size) to be as difficult as the Change problems, CA, and CA-2 (Hayes & Simon, 1977). In the current experiment, problem TA-2 was closer to problem TA in rule learning difficulty than to either of the Change problems. A subject-by-subject analysis of which rule sets were most rapidly learned supports the conclusion that the Move rules were easier than the Change rules: the rules for problem TA or TA-2 were each the easiest for four subjects, while the CA rules were easiest for none, and the CA-2 rules for only one.

Conclusion

The data on rule learning offer support to the **Rule Learning Hypothesis.** The long time it took subjects to learn the rules, with median learning times ranging from 1.5 min (Move) to 3.5 min (Change) (and individual times ranging up to 15 or more min) argues that differences in rule learning

[2] In addition, a group of three pilot subjects were given just two sets of rules (for the standard Move and Change problems, TA and CA'), and the number of repetitions in reading the rules was counted. Since this measure was highly correlated with the time measure, it was not used in subsequent analyses. For the three pilot subjects, the Change problem took far longer than the Move problem (394 to 85 s) and involved more repetitions of rules (43 to 19 repetitions), including the 6 repetitions demanded as evidence of reaching criterion.

TABLE 2
Rule Learning Difficulty

	Problem type				Avg.	Avg.
	TA	TA-2	CA	CA-2	TAs	CAs
Average time (s)	119	275	309	203	197	256
Median time (s)	97	88	223	178	94	206
No. times "easiest"	4	4	0	1	4	.5
Problem difficulty solution time (min)[a]	11.92	27.07	30.85	29.77	19.50	30.31

[a] From Hayes and Simon (1977).

difficulty may be a sizable contributor to problem difficulty differences (although they are not nearly large enough to be the sole cause). This conclusion is particularly clear for the pair of problems CA and TA, which had median rule learning times of 223 and 97 s, respectively, and a similar problem-difficulty ratio, generally about 2:1.

These differences in rule learning times would seem to conflict however, with reading times, which are approximately equal for the TA and CA′ problems. Analysis of the time subjects take to read the problem statements and rules, or the time until they begin to make moves (medians of 153 and 129 s for CA′ and TA, respectively) when compared with the rule learning times of 223 and 97 s, demonstrates that subjects start making moves, particularly on Change problems, before they have adequately learned the rules that govern moves.[3] We performed an experiment that investigated a number of features of this rule-using behavior.

EXPERIMENT 2: RULE APPLICATION

An experiment was conducted to investigate the subjects' ability to make judgments of move legality. A rule was presented tachistoscopically to the subject who studied it as long as she/he wished. Then a situation was presented representing a move that was either legal, illegal, or irrelevant under the rule. The subject was timed in responding with a judgment of the legality of the move.

[3] This conclusion is supported by the results of an analysis of the verbal protocols of a group of 16 subjects utilized in another experiment. One outcome of that analysis that is relevant here is their rule accessing behavior. For both Move and Change problems, there was a great deal of rule rereading after they started to make moves. This was especially true of the subjects solving a Change problem isomorph who reread a rule more often after starting to move than before, while the tendency for the Move problem subjects was the reverse ($\chi^2(1) = 4.78$, $p < .05$).

Method

Subjects. The subjects were 26 students at Carnegie–Mellon University who were paid to participate in the experiment.

Materials. The materials consisted of 96 cards with an upper field containing a typewritten rule and a lower field containing a pair of pictures. Typical stimulus cards are shown in Fig. 3. The rules were of four types patterned after Rule 2 or 3 of a Move or Change problem, with dots touching boxes substituted for monsters holding globes.

Thus while the wording was a bit different, the degree of restriction, the type of action (a change to a different sized entity, or a move of an entity that remains fixed in size), and the number of comparison loci involved, all were identical to Rule 2 or 3 of a Monster Move or Change problem. The pictures were of six types. They showed either two or three groups of objects and represented a move that was either legal (specifically permitted by the rule), illegal (specifically prohibited by the rule), or to which the rule did not apply. In Fig. 3, the left box of each example contains the *before* stimulus, and the right box contains the *after* stimulus.

Procedure. The subject viewed the cards in a two-field tachistoscope. The instructions were as follows:

> When you are ready for an experimental trial, say "ready." I will then expose the rule in the upper part of the screen. When you have read and understood the rule, press the center button. When you do this the rule will disappear and a pair of pictures representing a change will appear in the lower part of the screen. The situation before the change is represented on the left, and after the change on the right. If the change is legal according to the rule you just read, hit the button on the right; if illegal, the button on the left. Any questions?

Results

Reading times. Table 3 shows that reading times for Rule 3 are slower than for Rule 2 [$F(1,25) = 15.85$, $p < .001$] and are slower for Change than for Move problems [$F(1,25) = 16.81$, $p < .001$]. The average time to read Move problem Rule 2 is considerably faster than any of the other three rules. This rule, which involves comparing two things at the same locus, differs from the other three problem–rule combinations, which all involve comparing two items at two separate loci. The reading time for Change problem Rule 3, which is the longest, involves a doubly envisioned step—the comparison of a changed entity at a distance. We will have more to say about this spatial/imaginal property later.

Rule application times. The major dependent measure, move legality judgment time, is presented in Table 4 grouped by problem type, group size, and legality of move. The rules for Move problems are easier (more rapidly utilizable) than those for Change problems [$F(1,25) = 131.0$, $p < .001$]; within problem type, Rule 2 is easier than Rule 3 [$F(1,25) = 54.75$, $p < .001$];[4] and within rule and problem type, the two-group condition is

[4] There is a similarity here to one of the findings of Clark and Chase (1972) and Chase and Clark (1972) in their work on sentence–picture verification time. They found that sentences involving negation required more comparison time (608 ms in the sentence first

easier than the three-group condition [$F(1,25) = 109.5$, $p < .001$]. The correct response type (legal, illegal, does not apply) also yielded a significant effect; most notably, the does not apply condition was faster than either of the others in all conditions except Move problem Rule 2 [$F(2,50) = 40.24$, $p < .001$]. In addition, there were significant interactions, the largest being between group size and problem type, with the three-group condition being much harder than the two-group condition on Change problems, but only a little harder on Move problems [$F(1,25) = 28.15$, $p < .001$]. This is related to the structure of the two isomorphs in that the rules in the Move problem always entail tests at a single locus (Rule 2) or pair of loci (Rule 3). The third position (that not moved *from* or *to*) is irrelevant. In the Change problem, both rules potentially entail tests at all three loci. A similar effect is found for the interaction of group size and rule type, with the group size making more of a difference for Rule 3 than for Rule 2 [$F(1,25) = 9.64$, $p < .005$]. This finding of a multilocus positional difference between Move and Change problems (as well as between Rules 2 and 3) bears on the solution models discussed in Hayes and Simon (1977), and similar results appear with the memory load model discussed below. Finally, as with reading times, Move problem Rule 2 (the same-locus size-testing rule) was much faster than any other rule for either problem type, while Change problem Rule 3 was the slowest.[5]

Conclusion

The overall difficulty rankings for both the two-group and the three-group condition are, from easy to hard: Move problem Rule 2, Change problem Rule 2, Move problem Rule 3, Change problem Rule 3. This ranking is exactly the same as that obtained for rule reading times (Table 3). In Fig. 4 we depict the number of steps that must be envisioned in applying each of the rules, along with the rule application times from Table 4 for the three-group condition (the condition that most closely

condition) than the same sentence without negation. Similarly, Just and Carpenter (1976), in discussing their work on sentence–picture verification, conclude that since "pictures are generally represented affirmatively, so sentences that refer to pictures are generally easier to process if they are affirmative" (p. 207). There is such an affirmative–negative difference between Rule 2 and Rule 3 within each problem type. Rule 2 is an affirmatively worded rule while Rule 3 is a negatively worded one. This factor could therefore account for a portion (about half) of the additional Rule 3 comparison time. The differences found across the Move–Change boundary for similarly numbered rules, however, do not map onto the categories investigated by Clark and Chase, and therefore are in need of analysis, as are the unaccounted for portion of the Rule 2–Rule 3 reading time differences.

[5] The other significant interactions were between response condition and all other variables. The interactions were largely due to the fact that in the does not apply condition, the other variables have much smaller effects in all cases than they do in the legal and illegal conditions.

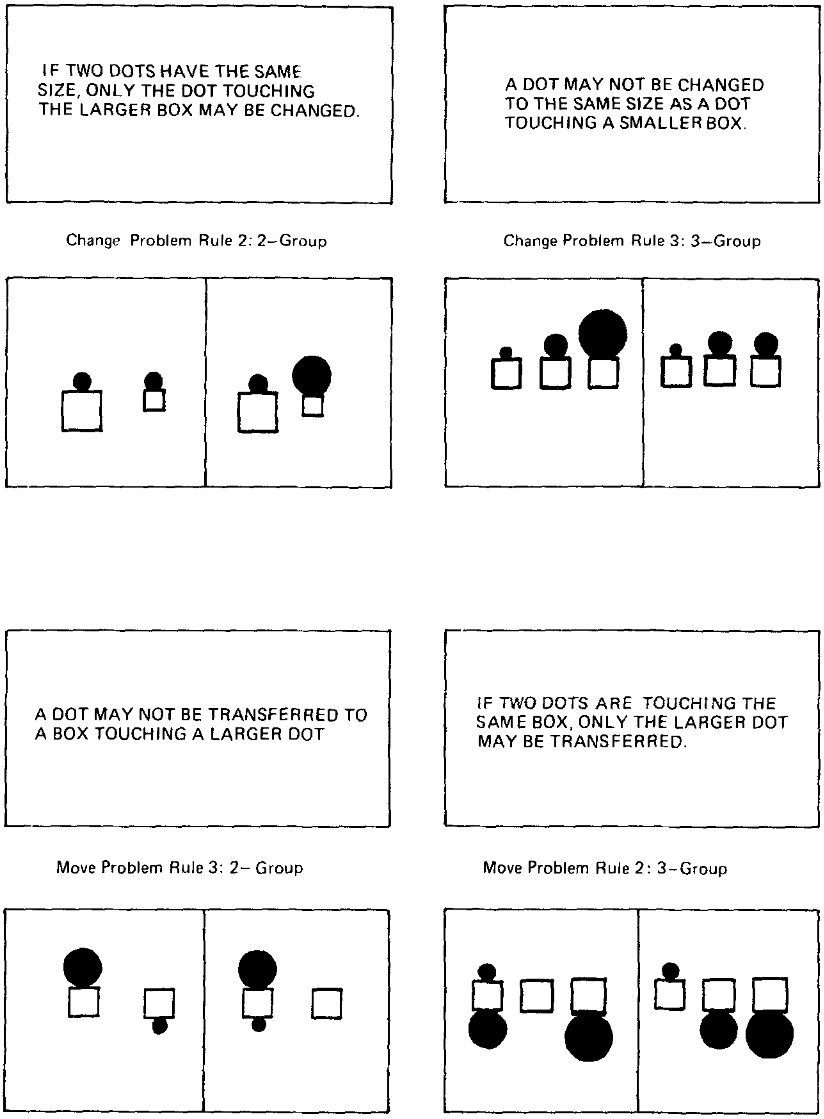

FIG. 3. Four examples of stimulus cards from Experiment 2: Rule Application. The top box containing the rule statement was exposed. Then when the subject was ready, she/he pushed a button replacing it with the lower panel containing the before (left box) and after (right box) display.

resembles the actual problem-solving situation). The number of envisioned steps is calculated by counting as a single step, either the imagined movement of an object from one location to another (to test size at a distance), or the imagined change of an object from one size to another

TABLE 3
Rule Application: Rule Reading Times (s)

Rule	Problem type: Move	Problem type: Change
2	6.83	8.18
3	9.12	9.72

(to test the equality of the new sizes). The fact that the size equality test itself requires a processing step is indicated in the first column of that figure. Given these definitions, Move problem Rule 2 involves no envisioned steps, Move problem Rule 3 involves one step—testing size at a distance, Change problem Rule 2 involves one step—testing size at a distance, and Change problem Rule 3 involves two steps—imagining a change in size, followed by size comparison at a distance. This simple model of the processing load is wholly consistent with the findings discussed above.

In summary, then, Rule 2 is easier to learn and use, especially on Move problems; the rules for Change problems are more difficult to learn and to apply than the rules for Move problems; adding extra positional information (a third group) greatly slows the process of rule application, particularly on Rule 3 and on Change problems, and these differences correlate with the memory load imposed by the representation of the various legality tests involved in making moves.

These effects, taken together with the previously discussed findings that subjects start to move before adequately understanding the rules, (and thus have to refer back to them frequently while making moves) argue that The **Rule Learning Hypothesis** is essentially correct in identifying rule learning as one determinant of problem difficulty. Moreover, difficulty of applying the rules also correlates with overall problem difficulty, as postulated by the **Rule Application Hypothesis.** If the learning of the rules and differences in rule application difficulty are important determinants of subject behavior on different problems, this suggests that similarity in rule structure might be an excellent predictor of transfer of training between problems.

Previous studies of transfer of training have focused on transfer between problem types that differed in difficulty (Move–Change), as well as those that differed in the semantics of their rule structures (Agent–Patient). Semantic differences had little effect, but the Move–Change dimension made a large difference. There was essentially no (or very little) transfer across the Move–Change boundary, while within each problem type, even across Agent–Patient differences, the amount of transfer was sizable, up to 50% (Hayes & Simon, 1977).

TABLE 4
Rule Application: Move Legality Judgment Time (s)

	Problem type			
	Move condition		Change condition	
	Two group	Three group	Two group	Three group
		Rule 2		
Legal	3.44	3.89	4.84	6.35
Illegal	3.65	4.15	4.82	6.22
D.A.	3.68	3.53	3.73	4.80
Rule 2 Avg.	3.59	3.86	4.46	5.79
		Rule 3		
Legal	6.42	7.12	6.30	9.72
Illegal	5.27	6.32	5.27	6.83
D.A.	3.80	4.23	4.44	4.87
Rule 3 Avg.	5.16	5.89	5.34	7.14

In the experiments reported below, we explore some effects of everyday knowledge on problem difficulty and transfer in Move problems. Everyday knowledge of the world might influence problem solution in a number of ways. Consider the Monster problem, TA, in Fig. 1 and the Acrobats problem depicted in Fig. 5. The cover story of the Monster problem evokes no prior knowledge that would suggest that the movement of globes among monsters should be controlled by globe size. In contrast, the cover story of the Acrobats problem does suggest a rationale for a relation between size and movement, based on our knowledge of human strength. The subjects' dependence on memory of arbitrary rules is thus sharply reduced.

In the above case, knowledge was applied to help represent information

	Test	Image	Image	Time
Move Rules				
Rule 2	X			3.86
Rule 3	X	X		5.89
Change Rules				
Rule 2	X	X		5.79
Rule 3	X	X	X	7.14

FIG. 4. Rule memory load: major demands. Each envisioned or imaged entity that must simultaneously be kept in mind is indicated by an X in an image column. The times are the number of seconds required to perform a legality test using that rule in Experiment 2.

Three circus acrobats developed an amazing routine in which they jumped to and from each other's shoulders to form human towers. The routine was quite spectacular because it was performed atop three very tall flag poles. It was made even more impressive because the acrobats were very different in size: the large acrobat weighed 700 pounds; the medium acrobat, 200 pounds; and the small acrobat, a mere 40 pounds. These differences forced them to follow these safety rules:

1. Only one acrobat may jump at a time.

2. Whenever two acrobats are on the same flag pole, one must be standing on the shoulders of the other.

3. An acrobat may not jump if someone is standing on his shoulders.

4. A bigger acrobat may not stand on the shoulders of the smaller acrobat.[a]

At the beginning of their act, the medium acrobat was on the left, the large acrobat in the middle, and the small acrobat was on the right. At the end of the act, they were arranged small, medium, and large from left to right. How did they manage to do this while obeying the safety rules?

FIG. 5. Acrobat and Reverse-Acrobat problems. [a]For the Reverse Acrobat problem, this rule was reversed, so that a smaller acrobat could not stand on a larger one, thus the large ones had freedom of movement in that version.

given explicitly in the problem statement. Everyday knowledge might also help us represent information that is implied by the problem statement but not stated explicitly. For example, moving in the Tower of Hanoi isomorphs involves manipulating lists in a last-in first-out manner, but this information is not given in either of the problem statements. It may be that the Acrobats problem helps subjects to make and represent this inference by eliciting our knowledge about flag poles and stacks of acrobats. A similar effect of *real world knowledge* on problem difficulty has been demonstrated in experiments on deductive reasoning in which the familiarity of the materials has a strong effect on subjects' performance (Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Shapiro, 1971).

EXPERIMENT 3A: REAL WORLD KNOWLEDGE AND RULE SIMILARITY

This experiment was conducted to measure the relative difficulties of the Acrobats problem and the Monster Move problem, TA, and the amount of transfer of training between them. The experiment tests the **Real World Knowledge Hypothesis,** that problems incorporating rule sets compatible with everyday world knowledge are easier than those which, though isomorphic in structure, have less compatible rules. This experiment is also a test of the **Rule Similarity Hypothesis,** that similarity between problem rule sets is a predictor of transfer of training between the problems.

Method

Subjects. The subjects were 20 Carnegie-Mellon University students who received course credit for their participation in the experiment.

Procedure. Ten subjects solved the Acrobats problem first and then solved the Monster problem, while the remaining 10 subjects solved the Monster problem first and then the Acrobats problem. Subjects were tested individually. The problems were presented individually on single sheets of paper in the format shown in Figs. 1 and 4. Subjects were allowed to use pencil and paper to aid solution. Solution time was the interval from the presentation of the problem until the subject achieved a correct solution.

Results

As Table 5 shows, the Acrobats problem takes less than half as long to solve as the Monster problem. The difference in solution time is significant by the Mann–Whitney U test at the .01 level. This problem was designed to be easy, the acrobat on flagpole scenario being chosen deliberately so as to provide a *natural* set of rules which would be easy to understand and remember. The results thus confirm the **Real World Knowledge Hypothesis.**[6]

Transfer of training in this experiment was moderate and not significant in either direction by Mann–Whitney U test ($p > .05$). Transfer from Acrobats to Monsters was −28.9% and from Monsters to Acrobats, 29.5%. This finding is inconsistent with an earlier result (Hayes & Simon, 1977) which found significant transfer between variants of Monster Move problems. Reasons for the difference are addressed after the presentation of the results of the second part of this experiment.

EXPERIMENT 3B: REVERSE ACROBAT STUDY

In this experiment, we explore further the hypothesis that the effects of the Acrobats cover story on problem difficulty can be attributed entirely to everyday knowledge. More specifically, we hypothesize that

1. The Acrobats problem is easy because the cover story, by incorporating real world knowledge, reduces the memory load required to remember the restriction of size on movement.
2. The lack of consistent transfer effects between Acrobat and Monster Move problems results from the lack of compatibility in the representation of the move operator restrictions.

To test these hypotheses, we constructed the Reversed-Acrobats problem (see Fig. 5 and its footnote) which was identical to the Acrobats problem except that the sense of the size-movement constraint was reversed and, therefore, made inconsistent with everyday knowledge, but

[6] A similar result was found for sentence understanding by Wason and Reich (1979), who presented subjects with sentences of the form, "No head injury is too trivial to be ignored," and "No missile is too small to be banned." They found that the latter sentence, which is more consistent with real world knowledge (*pragmatic knowledge*, as they termed it), was more often correctly understood, while sentences similar to the former were usually misconstrued in the direction of the pragmatic knowledge.

TABLE 5
Acrobat–Monster Move Problem Solution Time (min)

Problem type	Presentation order: 1st problem	Presentation order: 2nd problem	Amount of transfer
Acrobat	5.63	3.97	29.5%
Monster—TA	12.16	15.68	−28.9%

consistent with the form of the same rule in the Monster problem (i.e., in the Reversed-Acrobats and Monster problems, *large* restricts *small* from moving).

Method

Subjects. The subjects were 20 Carnegie–Mellon University students who received course credit for their participation in the experiment.

Procedure. Ten subjects solved the Reversed-Acrobats problem and then the Monster problem. The remaining 10 subjects solved the Monster problem and then the Reversed-Acrobats problem. In all other respects, the procedure was the same as in Experiment 3A.

Results

Table 6 shows that solution times for the Reversed-Acrobats problem did not differ significantly from the solution times for the Monster problem. The difference was not significant by Mann–Whitney *U* test ($p > .05$), whether the comparison was made just with the 10 subjects who solved the Monster problem first in Experiment 3B or with all 20 subjects who solved the Monster problem first in Experiments 3A and B. The difference in solution time between the Acrobats problem and the Reversed-Acrobats problem was also not significant ($p > .05$), although the ratio of the means was 1.7 to 1. Transfer effects in Experiment 3B were quite different from those in Experiment 3A. Transfer from Reversed Acrobats to Monster was +57.1% and was significant by Mann–Whitney *U* test ($p < .01$). Transfer from Monster to Reversed Acrobats was +38.5%, but was not significant by Mann–Whitney *U* test ($p < .05$).

TABLE 6
Reverse-Acrobat–Monster Move Problem Solution Times (min)

Problem type	Presentation order: 1st problem	Presentation order: 2nd problem	Amount of transfer
Reverse Acrobats	9.51	5.85	38.5%
Monster—TA	11.55	4.95	57.1%

Conclusion

The results are quite consistent with the hypotheses stated above. Unlike the Acrobats problem, the Reversed-Acrobats problem, in which the size-move restriction is no longer consistent with everyday knowledge, is not significantly faster than the Monster problem. However, transfer from Reversed Acrobats to Monster is very large—comparable to the amount of transfer among Monster Move problem isomorphs obtained by Hayes and Simon (1977). In this experiment and in the Hayes and Simon study, the sense of the size-move restrictions were the same in both problems—thus eliminating incompatibility, and allowing for large amounts of transfer.

Experiment 3 confirms the **Rule Similarity/Transfer Hypothesis,** (similarity of rule structure is a predictor of transfer of training). The data are also consistent with the **Rule Learning Hypothesis:** that the structure of the problem rules is a major locus of problem difficulty. Finally, using real world knowledge reduced problem difficulty significantly, thus confirming the **Real World Knowledge Hypothesis.** The next experiment investigates some of these issues further.

EXPERIMENT 4A: EXTERNAL MEMORY AIDS AND RULE TRAINING

This experiment was another test of the **Rule Application Hypothesis,** which posits that the tests to determine the legality of moves are an important source of differences in problem difficulty. Experiment 4A also tested the **External Memory Hypothesis** which states that differences in memory load may account for problem difficulty differences. One obvious difference between the Tower of Hanoi problem (which in its three-disk version takes only a minute or two to solve), and the other isomorphic problems is the external memory aid provided by the physical Tower of Hanoi, which accomplishes three things:

- It presents an easily accessed external memory for the current state of the problem.
- It removes from the subject the burden of generating a representation, and by presenting a more or less complete external representation of the problem prevents the subjects from using an idiosyncratic, less than optimal, representation.
- It embodies the rules in a perceptible, logical, and easily remembered manner. In particular, it virtually eliminates the need to remember Rule 2, which prohibits moving the larger of two disks when they are on the same peg. The physical placement of the smaller disk on top of the larger automatically prohibits movement of the larger. (This advantage was also present to an extent in the Acrobat problem because of the built-in Real World Knowledge.)

Experiment 4A employed a 2 × 2 × 2 factorial design in which the independent variables were (a) the presence or absence of physical models of monsters holding globes (which could actually be transferred or changed in size), (b) the administration or nonadministration or a training paradigm for rule application, and (c) problem type (Move or Change).

The first variable, presence or absence of monsters, represented an attempt to mimic some important aspects of the Tower of Hanoi problem by providing an easily accessible external memory for problem state. (The third feature of the disk–peg representation in the Tower problem, providing a logical and easily remembered mnemonic for Rule 2, was not realized.)

Method

Subjects. The subjects were 39 Carnegie–Mellon University students who received class credit for their participation. (One subject had to be removed from the analysis after the experiment had been run because it was discovered that he had made an illegal move that was mistakenly accepted by the experimenter.)

Procedure.

External memory and predefined representation. External memory was provided by three different sized papier-mâché monsters in which were embedded five pegboard hooks to serve as the five *hands* of the cover story and to hold the balloon *globes* of the problem. On Change problems, the subject changed the volume of a balloon by squeezing a clamp on an air hose connected to the balloon and blowing air into or releasing it from the balloon. The three air hoses were partially occluded so that the flow of air into or out of the balloons was easily controllable. For Move problems, the air hoses were not present. The balloons were simply moved back and forth between the monsters by means of wire loops that were easily hung on and removed from the monster hook *hands*.

RT training. The training was accomplished in the following manner. The subjects were told that they would be given some rules to learn, that they would be asked some questions about the rules, and that they should respond rapidly because they would be timed. A digital clock was in their field of view at this point, although about 60° off to one side so that it would not distract them during the actual trials. They were then given a sheet containing the three rules and asked to study the rules until they knew them. The subjects at this point had not been informed of the rules' role in any problem nor of the fact that they would be asked to solve a problem applying the rules. (The mean time taken by subjects to study the rules was 42.0 s and the standard deviation was 21.1 s.)

An initial situation consisting of an arrangement of monsters possessing globes was then presented to the subject and promptly represented, on the monsters or on paper according to condition. The subjects were each presented with five situations and asked seven questions about each.

An example of a Change problem question for the situation where the globe order is small monster–large globe, medium monster–medium globe, large monster–small globe, is "May the small monster make his globe medium?" This violates Rule 3 (because a larger monster already has a medium globe in that situation). The clock was started on the word *medium* and stopped when the subject responded "yes" or "no."

An example of a Move problem question for the situation where the globe order is small monster–large globe, medium monster–medium globe, large monster–small globe, is "Can the large monster give his globe to the medium monster?" This also violates Rule 3 since the medium monster already has a larger globe.

The clock was started on the key word (that is, on the earliest word in the question that allows for an unambiguous determination of legality or illegality of the move). The 35 questions involved one violation of Rule 1 (an easy rule for the subjects to learn and remember), eight violations of Rule 2, eight violations of Rule 3, one violation of both Rules 1 and 2, three violations of both Rules 2 and 3, and 14 nonviolations. By the end of the training, the subjects' responses were rapid and contained few errors. (Thus for the questions on the last situation, there were only four errors (1 Move, 3 Change) in a total of 140 responses, an error rate of less than 3%.) For the last half of the trials (the last 17 questions for each subject), the error rate was only 5%, and there was no difference in response times—the average time for this last half of the training trials was 1.73 s for the Move problem, 1.71 s for the Change problem. In contrast, the error rate on the 140 responses to the seven questions of the first situation (during the first portion of the training) was 27% for Change problems, and 13% for Move problems. The times for Change and Move problems on the first situation were 2.77 and 1.77 s, respectively. After the training, the subjects were given their problem (either Move or Change) in the same manner as subjects who had received no rule application training.

Problem solving. The subjects were then given a sheet of paper with the problem on it (as well as paper and pencil) and asked to start, and the clock and recorder were started. They were expected (and prompted if necessary) to read the problem aloud. Subjects who received no training were asked to set the monsters up with globes of certain sizes initially (prior to the start of the clock) so that they obtained experience with the balloon system.[7]

The dependent measures were the number of moves until a solution was reached, the time it took to solve the problem, the time to read the problem, the time until the first move, and the time at which each move was made (this latter measure was taken during the solution attempt for some subjects and from the tape recording for others).

Results

The major dependent measure was solution time, logarithmically transformed for the ANOVA. The results for this measure are presented in Table 7 and in Fig. 6. Problem type—Move versus Change, the presence of the physical models of monsters, and RT training—were all significant. The Move problems were easier [$F(1,31) = 4.6$, $p < .05$], the presence of the physical monsters made the problems easier [$F(1,31) = 5.8$ $p < .05$], and the training was effective at reducing problem difficulty [$F(1,31) = 11.6$ $p < .005$]. Interestingly, a second measure of problem difficulty, number of moves made during the solution attempt (depicted in Table 8), yielded no effect of training nor of problem type, but did show an effect for monsters, with the presence of monsters tending to decrease the number of moves [$F(1,31) = 7.2$, $p < .05$]. These two findings taken

[7] Because of the external memory condition, the order in which the monsters were mentioned in the problems was changed from that used in previous work so that they were mentioned according to increasing size. In earlier work, the order of mention was deliberately permuted (it was medium, large, small); in this study it was small, medium, large. This was done to eliminate extraneous differences that might arise from the placement of the physical monsters for subjects run in that condition of the experiment. The monsters had to be placed in some order, and size order was chosen, since most people arranged their paper and pencil representations that way in earlier versions of the experiment no matter what the order of mention in the problem statement.

TABLE 7
External Memory Aid–Rule Training Experiment: Average Solution Time (min)

	Problem type						
	Move			Change			
	RT trained	Not trained	Move Avg.	RT trained	Not trained	Change avg.	Grand avg.
Monsters present	4.47	6.12	5.30	5.43*	19.17	12.30	8.80
No monsters present	7.70	14.08	10.90	10.17	13.57	11.87	11.38
Average	6.08	10.10	8.10	7.80	16.03	12.08	10.10

* $n = 4$.

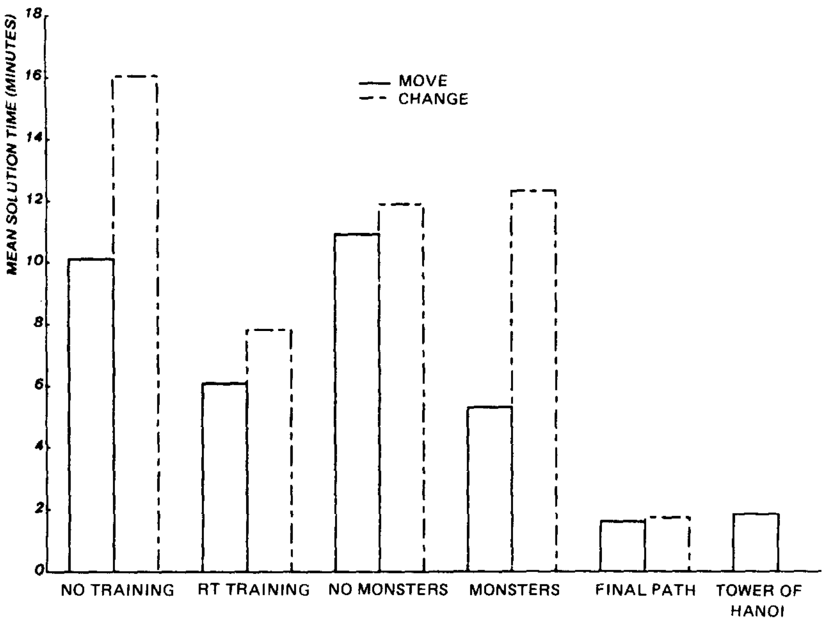

FIG. 6. Isomorph solution times and final path times from Experiment 4A, and Tower of Hanoi solution time.

together demonstrate that training and problem type affect the time per move, rather than the number of moves. The largest effect found was that of RT training on solution time, which held across each problem type and each monster condition, as can be seen in Table 7.

The data on errors are presented in Table 9. The average number of errors was 2.2 per subject. There are almost no violations of Rule 1, while violations of Rule 3 outnumber violations of Rule 2 in all cases *except* for a Change problem with monsters absent (MA), where Rule 2 violations were relatively abundant. Neither RT training nor problem type had a significant effect on overall errors. As depicted in Fig. 4, application of Change problem Rules 2 and 3 and Move problem Rule 3 all require comparison of the sizes of globes (and/or monsters) at two different loci, while Move problem Rule 2 entails comparing the sizes of two things at the same locus. Table 9 shows that the least errors were committed in the Move problem single-locus condition (Rule 2), which Experiment 2 showed to be the most rapidly read and the most rapidly applied rule.

The presence of the physical monsters (MP) did reduce errors [$F(1,31) = 4.9$ $p < .05$], but, interestingly, a large part of this effect consisted of errors that were made but then caught by the subjects in the 2 to 3 s preceding the experimenter's intervening. There were 12 of these *self-*

TABLE 8
External Memory Aid–Rule Training Experiment: No. of Moves/Problem

	Problem type						
	Move			Change			
	RT trained	Not trained	Move avg.	RT trained	Not trained	Change avg.	Grand avg.
Monsters present	11.2	11.2	11.2	10.3*	17.4	14.2	12.7
No monsters present	17.4	18.2	17.8	16.8	16.8	16.8	17.3
Average	14.3	14.7	14.5	13.9	17.1	15.5	15.0

* $n = 4$.

TABLE 9
External Memory Aid–Rule Training Experiment: No. of Errors

Move—monsters	0	5	8	13
Move—no monsters	1	6(9)	13(17)	20(27)
Change—monsters	0	5	9	14
Change—no monsters	0	17(18)	8(12)	25(30)
Σ	1	33(37)	38(46)	72(84)

Note. The numbers in parentheses include rule violations that were made but then noticed by the subject before the experimenter intervened. They are termed "self-caught" errors in the text.

caught errors (out of a total of 84 errors), all 12 occurring in the MA condition. (If these *self-caught* errors are removed from the analysis, then the effect of MP on errors no longer reaches significance.)

A related finding is that subjects at times stop their step-by-step movement through the problem space and start over from the beginning (23 instances) or from some prior position in the problem space (1 instance). Of the 24 instances, 23 occurred in the MA condition; thus, self-caught errors and starting over behavior both occur almost exclusively among the subjects who are using paper and pencil rather than monster model external representations.

Conclusion

The two hypotheses tested simultaneously in Experiment 4A are the **Rule Application Hypothesis** and the **External Memory Hypothesis.** The rule training did reduce problem difficulty significantly as did the presence of physical monsters (MP). Also, while the rule training tended to reduce the difficulty differences between Move and Change problem, the MP condition did not, and neither condition eliminated the differences between the Tower of Hanoi problem and the Monster problems.

When we discussed the rationale for the **External Memory Hypothesis,** we identified three potentially facilitating effects of the Tower of Hanoi disk–peg representation: (1) the presence of an easily accessible external memory for problem states, (2) the removal from the subject of the burden of creating a representation, and (3) the provision of a logical and easily remembered mnemonic for the problem rules governing moves, which virtually eliminated the memory load imposed by Rule 2.

The external memory aid provided by the physical monsters (MP) addressed the first two of those features, but not the third. Thus, the failure of the MP condition to eliminate the difficulty differences between the Move and Change problems (and to eliminate the even larger differences between both of these and the Tower of Hanoi problem) again suggests the problem rules and their application as the source of differences in difficulty. The finding that rule training significantly reduced problem difficulty across all the other conditions of the experiment and tended to reduce the difficult differences between Move and Change problems supports this view, although the failure of the rule training to wholly eliminate the Move–Change differences (it only reduced the difference from 59 to 29%) weighs against it.[8]

EXPERIMENT 4B: EXTERNAL MEMORY AIDS: TOWER/DISH PROBLEM

In order to clarify further the role of external memory aids in reducing problem difficulty, an additional experiment was performed. The problems were isomorphic examples of the three-disk, five-step Tower Disk–Peg problem that varied in the degree to which the rules governing moves were built into the external representation provided.

Method

Subjects. The subjects were 45 students at Carnegie–Mellon University who were given course credit for their participation.

Procedure. Each subject was given two problems, one at a time, of which we will use only the first.[9] The subjects were instructed to solve the problem while thinking out loud as in previous experiments. They were tape-recorded and timed as they solved the problem, and the experimenter was present to intervene in case of an illegal move that the subject did not catch, and to prompt for verbalization.

Materials. There were three problems that differed in the degree to which they incorporated problem rule information into the external physical representation of the problem. The first was the Peg–Move problem which was nearly identical to the standard Disk–Peg problem. The only difference was that the apparatus consisted of three dishes (ceramic flower pot bases of different sizes) each containing a vertical peg (wooden dowel rod). Instead of the more usual flat disks, three styrofoam balls of varying diameter were impaled on the pegs. The balls were 1.5, 2, and 2.5 in. in diameter, and each had a hole drilled through its center so that it could easily slide on and off the pegs. The subjects' goal was to move the balls from peg to peg until the size order of the balls matched the size order of the dishes. This isomorph thus incorporated in the external physical problem representation the Rule 2 constraint that only the smaller of the balls on a peg may be transferred.

The second problem isomorph was termed the "Dish–Move problem." It was identical to the Peg–Move problem with the following change. There were no pegs. The balls were instead placed in one or another of three dishes. The dishes were of three sizes, and the

[8] This might be due to the unusual results obtained from one cell of the experiment. The Change problem, no RT, MA subjects (those most similar to subjects in previous experiments), exhibited faster solution times than any prior group of subjects on that problem. The comparison of those subjects with subjects in the RT condition might thus be misleading in indicating a lack of effect of RT for this condition; the RT group was being compared with an unusually fast group.

[9] This experiment used a repeated measures design to determine the amount of transfer from the first of two problems solved to the second. The effects to be discussed here are for the first problem only.

subject's task was to move the balls from dish to dish until the order of ball size matched the order of dish size. This Dish–Move isomorph, while providing a good external physical representation of many aspects of the problem (in the same manner as the physical monsters of Experiment 4A), did not incorporate the Rule 2 constraint in the physical representation as did the Peg–Move problem.

The third problem, termed the Dish–Change problem, not only did not incorporate Rule 2 information in the external representation, but also required more size comparisons in making moves. The problem utilized the same three dishes (without pegs), but had in addition, three *reserve dishes*. Each reserve dish was placed directly behind a problem dish. The subject's goal was the same as in the other two problems, to end up with each problem dish containing the appropriate sized ball. The difference was that this problem only allowed exchanges between a dish and its reserve dish. For example, if a subject wanted to change the ball in Dish 2 from medium to small, she/he had to remove the medium-sized ball from Dish 2 (placing it in Dish 2's reserve dish), and take the small ball from Dish 2's reserve dish and put it into Dish 2. This problem is labeled a ***Change*** problem because the moves are like those of the other Change problems in that they occur in only one locus (one size interchange is made at a time) and require equality tests with the ball sizes at each other location for both Rules 2 and 3. However, since balls are actually interchanged between a dish and its reserve dish, nothing really changes in size. Therefore, the Rule 3 comparisons can be made between actual balls and do not impose two simultaneously envisioned steps, as in the usual Change problem (Fig. 4), thus eliminating what we argue to be a major source of more usual Change problem difficulty. The amount of information about the rules that is built into the physical problem representation is, from most to least: Peg–Move problem, Dish–Move problem, Dish–Change problem.

Results

On the basis of our previous analysis, the relative difficulties of the problems in the current study should reflect the relative memory loads imposed by the move operators for the three problems. The expected problem difficulty order is, therefore, from least to most difficult: Peg–Move, Dish–Move, Dish–Change. The results are shown in Table 10.

The observed problem difficulty order was the expected one. An analysis of variance of the log transformed times shows that the above result should be viewed as tentative. The ANOVA only approached significance [$F(2,42) = 2.70, p = .08$].[10] *T* tests on the individual means, which should be interpreted with caution, show that the hypothesis that the Dish–Move problem was harder than the Peg–Move problem had a t of 1.92 (df = 28, $p < .05$); the hypothesis that the Dish–Change problem was harder than the Peg–Move problem had a t of 2.19 (df = 28, $p < .025$); and the hypothesis that the Dish–Change problem was harder than the Dish–Move problem had a t of .35 (df = 28, $p < .1$). Thus, there is evidence that the difficulty order found is that which was predicted, with the possible exception of the Dish–Move/Dish–Change difference. In addition, the difference found between the Peg–Move and the Dish–Move prob-

[10] If the results from the entire transfer study (the first and second problems) are combined, then the main effect is significant [$F(2,42) = 4.09, p < .05$.]

TABLE 10
Dish Problem Results

Problem type	Solution time[a] (s)
Peg–Move	173.4
Dish–Move	255.3
Dish–Change	276.6

[a] The times reported are the antilog of the average of the log-transformed times. The nontransformed times show a similar pattern with the exception that the Dish–Move/Dish–Change difference is a bit larger.

lems confirms the interpretation reached in Experiment 4A, that the difference between the solution times for the isomorph with the physical monsters and the Tower problem was due to the fact that the former did not eliminate the memory load imposed by Rule 2 as did the latter. That difference in the representation of Rule 2 (whether it is built into the physical representation of the problem or not) is precisely the difference between the Peg–Move problem (which incorporates it) and the Dish–Move problem (which does not), and the result is the same: a difference in problem difficulty favoring the problem representation that incorporates the rule information.

DISCUSSION: FINAL PATH OF SOLUTION; THE TEMPORAL PATTERN OF MOVES

An unanticipated finding of Experiment 4A was the emergence of a temporal patterning of the subjects' moves. Once this pattern was noticed, early in the study, the time at which each move was made was recorded. For subjects who had already been run (and for those occasional subsequent subjects whose times were missed), the times were obtained from the tape recordings that had been made of the problem-solving sessions. Only one subject out of the 39 could not be timed adequately.

The behavior pattern was this: Subjects spent a large portion of their time in what was seemingly exploratory behavior, ultimately arriving back in the beginning state (five moves from the goal state). They then moved very rapidly and successfully toward the goal. This rapid closing in on the goal will be termed *final path behavior*. The temporal structure of the behavior is evident in the data presented in Table 11.

The data show that (1) the time spent on the final path constitutes only 15% of the total problem solving time (or 21% of the time spent on moves); (2) of the total number of moves, 38% are on the final path; and (3) the median time on the final path is slightly less than a minute and a half (77 and 95 s, respectively, for Move and Change problems, or 79 and 73 s, respectively, for the two-thirds of the subjects exhibiting a standard or near-standard final pathway). This similarity of the final path times for Move problems and Change problems, together with their brevity, is evidence that the significant problem-solving activity is essentially over before the subject starts to move toward a solution. As the subjects make their first move on the final path toward the goal, they are as far from it as they were at the beginning of the problem (five steps), yet they will attain the solution in slightly under a minute and a half. These times are approximately equal for Move and Change problems, and are also approximately equal to the time it takes subjects to solve the isomorphic Tower of Hanoi problem (Fig. 6).

These findings suggest that subjects are first exploring the parameters of the problem, learning how to move in the problem space, and learning the rules that restrict those moves, until they *see* a solution and proceed to execute it. While this interpretation is consistent with the findings presented so far, further analysis shows it to be too simple.

"Two-Chunk" Temporal Pattern

When the times of the moves are examined in detail, not only is the exploratory phase–final solution dichotomy found, but another pattern as well. For almost two-thirds of the subjects, the final path was a standard one or very close to it, and for over half of those subjects, the timing of moves within that final solution path exhibited a common and extremely interesting subpattern. The standard final pathway consists of the five moves minimally necessary to reach the goal from the initial starting point. Fifteen of thirty-nine subjects followed this five-step pathway. Another 9 subjects followed this pathway with minor variation (for 8 of the 9, an extra step on the third move, and for 1, an extra step on the second move). The remaining 15 subjects followed idiosyncratic pathways, introduced more than one extra move, or made illegal moves in approaching the goal.

Many of the subjects seemed suddenly to be able to plan two moves, would make these two moves, and would then plan the remaining three moves relatively quickly. This *two-chunk* process seemed to be characteristic of the final path behavior. This idea was tested in two ways. The subjects who were run after the pattern was discovered (27) were queried as soon as they completed the problem, as to how much of a plan they had when they made the first move on the final path (the particular move was pointed out to them). With few exceptions, they answered that they did *not* have the whole solution in mind at that point, but rather, at best, a partial goal or insight. Examples of subjects' comments were "My goal was to change small monster to a small globe. Large monster was free

TABLE 11
Final Path Behavior

	Problem type			
	Move		Change	
	All subjects	Standard and near-standard final path subjects	All subjects	Standard and near-standard final path subjects
Median final path time (s)	77	79	95	73
Median time/move on final path(s)	15	16	14	12
Median time/moves prior to final path (s)	23	26	31	34
Median (final path time as % of total time)	15	14	15	17
Median (final path time as % of move time)	20	19	22	22
% of total moves on final path	40	30	34	37
Median time to first move (s)	129	126	152	133
Final path errors as % of total errors	5	0[a]	11	0[a]

[a] An illegal move in the final path placed that subject in a category other than standard or near-standard. There were only two Move problem and three Change problem subjects who made illegal moves on otherwise near standard pathways. There were no errors committed by subjects on otherwise standard pathways.

to change. *Then* I saw that I could move large monster to small globe and do it'' or ''When I moved the small (globe) over, *then* I realized I could move the large over there (to the small monster) and thus move the medium.''

The verbal behavior of the subjects is thus consistent with the conclusion that they did not have the whole five steps to the solution planned when they began their rapid movement along the final pathway. The hypothesis that they moved in two chunks of two and three moves, respectively, was formally tested (using a technique developed by Chase & Simon (1973)) from the temporal pattern of moves. If subjects did indeed move in two stages to a solution, then the first move in each chunk should

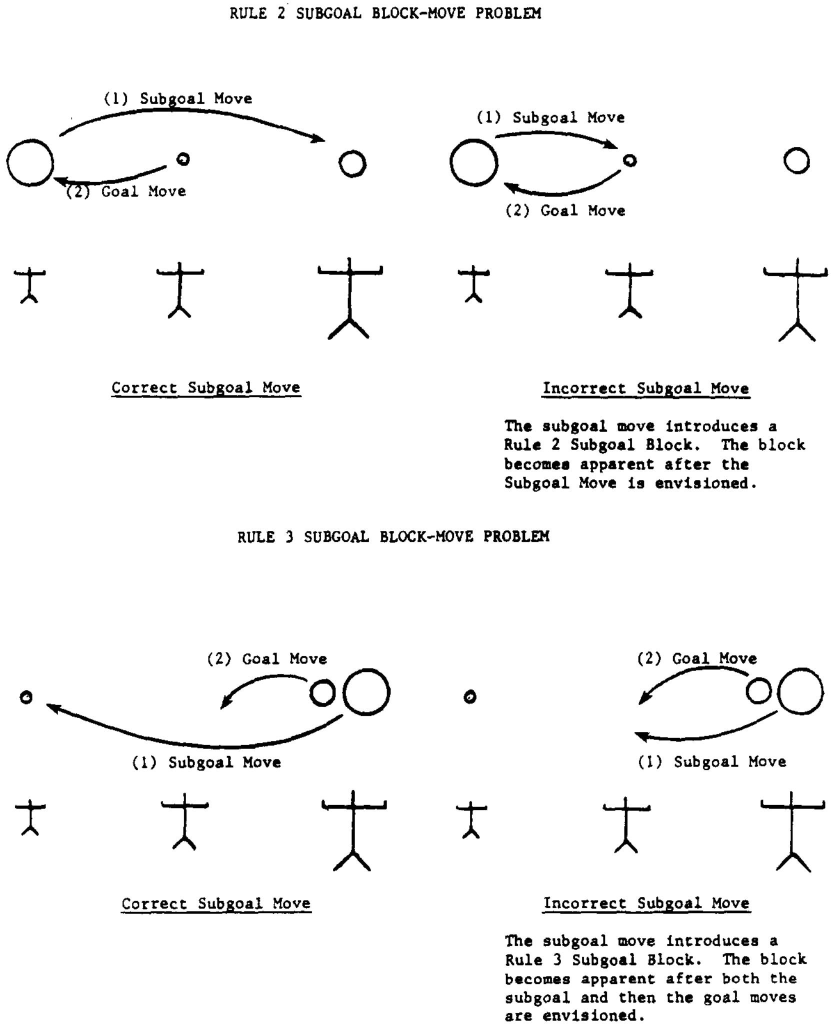

FIG. 7. Rule 2 Block and Rule 3 Block examples: Move problem. Correct (left side) and incorrect (right side) subgoal moves are depicted for situations in which the incorrect subgoal move would introduce a Rule 2 Block (top), or a Rule 3 Block (bottom), of a subsequent goal move.

take longer than the subsequent move or moves. This should lead to a pattern of times for the last five moves of long–short (Chunk 1), long–short–short (Chunk 2). The probability of this pattern occurring by chance is 1 in 10. A similar analysis for the near-standard six-step final path solutions yields three acceptable possibilities out of a total of 15 arrangements, for an expected chance occurrence of 1 in 5. Of the 15 subjects exhibiting a standard final pathway, 10 exhibited the above pattern, 4 exhibited temporal patterns inconsistent with this one, and the remaining subject was marginally in accord with it. Among 9 subjects exhibiting final pathways that were close to standard, 4 exhibited the two-chunk temporal pattern, 2 were inconsistent, 1 marginally fit, and 2 were not susceptible to analysis because their first move on the final pathway was their first move in the problem. Thus, while the number of move patterns expected to fit the two-chunk hypothesis by chance was approximately 3; the number observed was 14–16. This constitutes strong support for the hypothesis ($p < .0000005$ for standard final path subjects, and $p < .05$ for near-standard final path subjects, binomial test—treating the marginal cases as negative outcomes).

The conclusion that planning one move ahead (compiling two moves) is the key factor in beginning the final path sequence is not only supported by the two-chunk pattern and the verbal reports, but also by a detailed analysis of the move patterns and latencies for the moves of both Chunk 1 and Chunk 2.

Subgoal Planning

The moves that make up the two chunks were examined further to determine the factors that account for the temporal pattern, and to account for the nonoptimal solution path taken by the near-standard-path subjects.

The first chunk involved the formation and execution of a subgoal–goal strategy to get the small monster the correct globe. In the Move problem, TA, this entails moving the large globe from the small monster to the large monster so that the small globe can then be moved to the small monster. In the Change problem, CA′, the large monster's globe is changed from small to large so that the small monster's globe can be changed from medium to small. These *unblocking* first moves both involve executing a subgoal so that a higher level goal can be accomplished. This is illustrated in Fig. 7 for the Move problem. This pair of moves is termed a "Rule 2 Block" situation because an incorrect move (of the large globe to the wrong monster/size) would introduce a block of the goal move via Rule 2, as is depicted in Fig. 7.

The second chunk also consisted of a subgoal–goal pair of moves, followed by a final move of the last globe to its final state. After the completion of the first chunk for Move problem, TA, the small monster has the small globe, the medium monster holds nothing, and the large monster has the medium and large globes. (This situation is depicted in the "Rule 3 Block" portion of Fig. 7.) The medium globe cannot be moved to the medium monster because it is blocked (via Rule 2) by the large globe. The *standard path* (and shortest) solution path is to move the large globe to the small monster, the medium globe to the medium monster, and the large globe back to the large monster. The near-standard-path subjects, however, adopted the goal of moving the medium globe to the medium monster and the subgoal of moving the large globe so as to *unblock* the goal move, but could not plan far enough ahead to recognize that a wrong subgoal move (large globe to medium monster) will block the subsequent goal move (medium globe to medium monster) via Rule 3. This is designated a "Rule 3 Block" situation. Having made this wrong move at Step 3, the subjects then have to introduce an extra move to remove the error.[11] The final step, which requires no subgoal processing, is then executed very rapidly. The Change problem, CA′, is susceptible to exactly the same analysis.

That the second chunk pair of moves is a subgoal–goal pair is evidenced by the temporal move patterns of 14 out of 15 standard path subjects, and 7 out of 9 near-standard-path subjects. The probability of this occurring by chance is $p < .001$ (binomial test).

Finally, under what conditions can subjects compile such two-move, subgoal–goal move sequences? While most subjects did do so in both the first and second move chunks, 8 of the 9 near-standard subjects did make the wrong move on Step 3, and 1 on Step 1. We argue above that they did so because they cannot completely compile the two-move sequence, even though (for at least 6 of the 9) they had successfully compiled the two-move first chunk subgoal–goal pair. What determines whether or not the subjects can plan and execute complete two-move sequences in the final paths of these problems?

The data are summarized in Table 12, where it can be seen that in 31 of the 32 occurrences of a Rule 2 Block, the correct subgoal–goal move pattern was executed. Among 25 move pairs involving a potential Rule 3 Block, the correct move was chosen in only 17. This difference between the Rule 2 and the Rule 3 block was significant [$\chi^2(1) = 6.76$, $p < .01$].

Table 12 also contains the results of a similar analysis of the final path moves of the remaining 15 subjects (those who were classified as neither standard nor near-standard). The final paths for these subjects were defined as the moves made after their last occupancy of a position five steps away from the goal. There were 23 instances of a potential Rule 2 Block, and in 20 of them, the correct move was made. There were 22 instances of a potential Rule 3 Block, and the correct move was made in only 10 of them. Again, the Rule 3 Block situation is harder [$\chi^2(1) = 6.95$, $p < .01$]. However, evidence that the subjects were indeed engaging in subgoal–goal planning behavior is much stronger for the first group of subjects than it is for the second.

[11] An unexpectedly high error rate at the beginning of Chunk 2 was also discernible in the behavior of a prior group of 13 subjects who unexpectedly made more errors when three moves from the goal than when four moves away.

TABLE 12
Subgoal Move Behavior

	Standard & near-standard subjects (n = 24)				Nonstandard subjects (n = 15)			
	Move		Change		Move		Change	
Type of subgoal block:	Rule 2	Rule 3	Rule 2	Rule 3	Rule 2	Rule 3	Rule 2	Rule 3
No. correct moves	19	7	12	10	11	4	9	6
No. incorrect moves	0	6	1	2	3	7	0	5
No. illegal moves	0	0	0	0	0	2	1	4

The greater difficulty that both subject groups had choosing the correct move in the Rule 3 Block situation is attributable to the necessity of imaging (in some form) both the subgoal move and the goal move, and then noticing (if the wrong subgoal move is chosen) that the goal move is blocked by the potential Rule 3 Block. The Rule 2 Block situation involves only imaging the subgoal move and noticing (if the wrong subgoal move is chosen), that the goal move is blocked by Rule 2. The difference is in comparing an imaged entity with a real one, as opposed to comparing an imaged entity with another imaged one. Figure 8 portrays the number of envisaged steps that must simultaneously be kept in mind in the two situations.

A RANDOM WALK ANALYSIS OF THE SIZE OF THE PROBLEM SPACE

The first half of the pattern of behavior discussed above (exploration followed by rapid movement to the goal) was further compared with a random walk through the problem space. Starting at the initial problem state, moves were selected randomly, with all legal nonrepetitive moves

	Test	Image	Image	Image
Move Planning				
Rule 2 Goal	X	X		
Rule 3 Goal	X	X	X	
Change Planning				
Rule 2 Goal	X	X	X	
Rule 3 Goal	X	X	X	X

FIG. 8. Subgoal memory load: major demands. Each envisioned or imaged entity that must simultaneously be kept in mind in planning a Rule 2 Block or Rule 3 Block Subgoal–Goal move pair is indicated by an X in an image column.

from each position equally probable, until the goal, which was five moves from the starting position, was reached.

The model had two parameters: (a) the probability of returning to the immediately preceding position and (b) the distance from the goal at which the walk was halted. The first parameter was set at .03 to match the average probability of subjects undoing their previous move. Figure 9 shows the number of moves required by the model to get to the goal or within one, two, or three moves' distance from it, respectively. The figure also shows the average number of moves that subjects took to reach these targets.

The figure shows that the number of moves the random walk model took to get within three steps of the goal was similar to the number of moves that subjects took to reach the same position, but that subjects reached the goal with an average of 5 additional moves (for 20 of the 39 subjects, in the minimum 3 moves), while it took 23 additional moves for the random walk. The percentage of correct moves made at various distances from the goal by a previous group of 13 subjects were 27, 82, 68, 100, 100 at 5, 4, 3, 2, and 1 steps' distance from the goal, respectively. The results were nearly the same for the Move and Change problems. Each point in Fig. 9 represents the results of 1000 random walks. Thus, a random walk may approximate the exploratory behavior of subjects, but it does not model their final path behavior.

EXPERIMENT 5A: HINTS

The dichotomous nature of the subjects' behavior could result from either their difficulty in planning ahead, or their lack of knowledge of the direction to take in order to reach a point from which the problem becomes easy. In order to discriminate between these possibilities, a *hint* study was undertaken that was designed to enable the subjects to make the first two moves and thus be well along the final pathway. If the difficulty that occurs in the exploratory phase is caused by lack of knowledge about the direction that leads toward the goal, then a strong statement informing the subjects about that direction should be immensely useful in putting them into the second phase of the problem-solving behavior. If, on the other hand, the difficulty lies in subjects' inability to compile two moves and thus to execute the first chunk of the standard final path, then the hint should not be particularly helpful.[12]

[12] Another possibility is that the exploratory phase of the subjects' behavior simply familiarizes them with the portion of the problem space that must be traversed on the way to the goal. Examination of the moves the subjects make, however, dispels this idea. As the data discussed above on moves versus goal distance show, once subjects get to within three (or usually), even four, moves of the goal, they are on the path to a solution. Subjects penetrate very little of the final path during their exploratory behavior, and thus their acquisition of the ability to quickly move down the final path is not explained by their gaining knowledge of the particular moves that lead to the goal.

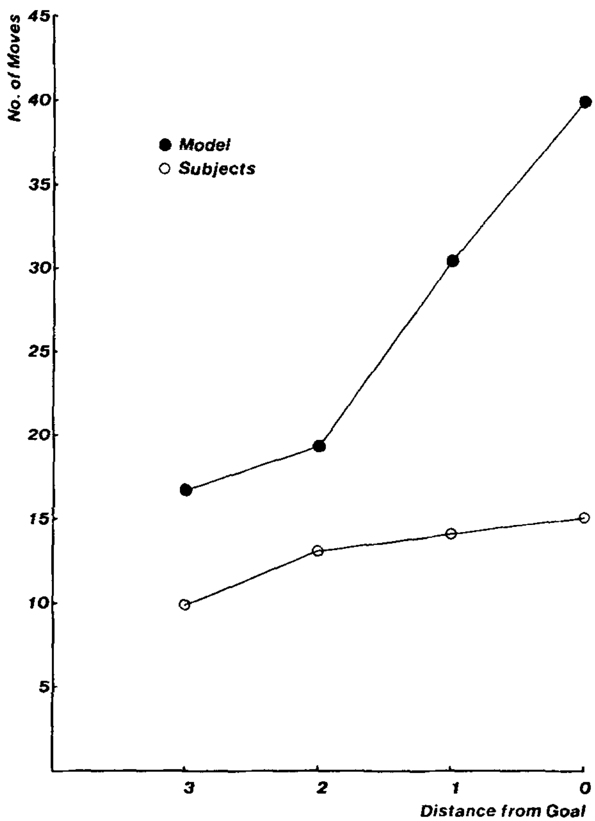

FIG. 9. Move data: random walk model and human subjects. While the model approximates the number of moves required to move part way toward the goal, the subjects reach the goal with a small number of additional moves, while the model requires many more.

Method

Subjects. The subjects were 11 Carnegie–Mellon University students who received course credit for their participation.

Procedure. The methods used in this study are equivalent to those of Experiment 4A with the following changes.

- There were no physical monsters, and no RT training.
- At the bottom of the instruction sheet, a sentence was added which constituted the strong hint, informing the subject what to do first. It is seen in brackets at the bottom of Fig. 1 for problems TA and CA′. If the subjects did not solve this problem in 5 min, they were asked some time soon after that about their current goal, and if they did not express the hint as their goal (to get the small monster the right globe), it was repeated orally to them.
- The subjects were given a second problem to solve upon completion of the first, so that each subject solved a Move and a Change problem. Only the first problem was used in the analysis to allow the times to be compared with the data from corresponding conditions (no training, no physical monsters) of Experiment 4A.

Results

The results are presented in Fig. 10, together with the data from corresponding conditions of Experiment 4A. While there is a small tendency for the hint to decrease the solution time in the Move problem, it is not significant [$t(4) = 1.4, p > .05$] and the tendency for the Change problem in the opposite direction is also nonsignificant [$t(5) = .9, p > .05$]. (The only subject to fail to solve a problem in the experiments was a single subject working on the Change problem in the hint condition.) Overall, there are no significant differences at 5% level between the hint and no-hint conditions in moves, total solution times, or move times (Wilcoxon signed ranks test, or t test, where appropriate). The differences between the hint and final path times, however, were significant, with the final path times (from Experiment 4A) being faster than the hint times on both Move and Change problems [$t(4) = 3.77, p < .001$] and [$t(5) = 4.76, p < .001$], respectively. The hint certainly did not have the effect that was being tested: that of putting the subjects onto the final pathway.

EXPERIMENT 5B: HINT TOWER OF HANOI STUDY

As a further test, a similar hint study was run using the isomorphic Tower of Hanoi problem.

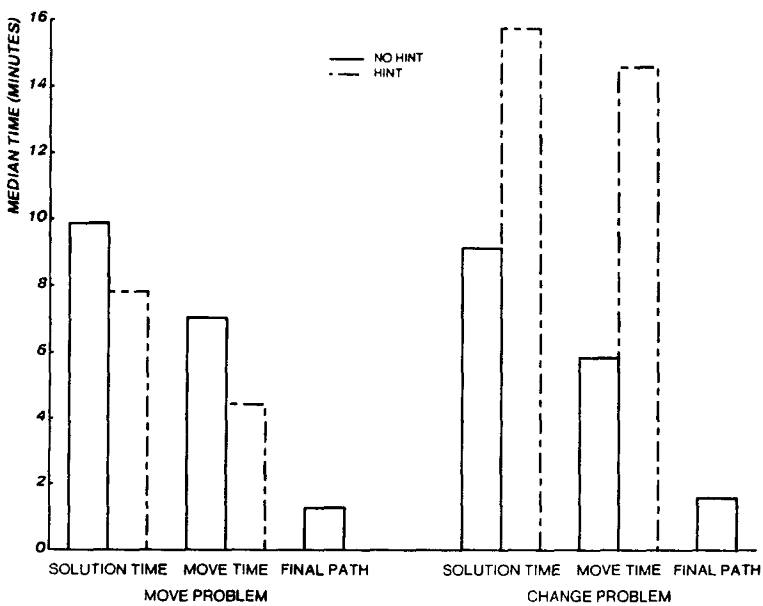

FIG. 10. Monster Problem hints: Experiment 5A. Solution times and move times for the hint condition are compared with equivalent (but no-hint) subjects from Experiment 4A and with the final path times from that experiment.

Method

Subjects. The subjects were 30 community college students who were paid for their participation.

Procedure. The subjects were given a Tower of Hanoi problem consisting of three disks on three pegs. In order to make the problem an isomorph of the monster problems, each disk was placed on a separate peg, thus making this a five-step rather than the more usual seven-step problem. Subjects were instructed to "move the disks so that each disk would end up on its own colored peg" (there were three colors of disk and a matching three colors of peg). There were two conditions: a nonhint condition, and a hint condition in which the subject was also told to "take care of the white disk first" (i.e., get it on its right colored peg). As above, this hint, if followed, required the subject to move one disk out of the way in order to get the white one on its peg, thus accomplishing the first two of the five required moves.

Results

The results are presented in Fig. 11 where it can be seen that while there was not a *total* solution time difference, there was a difference in *move* times favoring the hint [$t(14) = 2.19, p < .025$]. Thus, in problems where the subjects have a physical embodiment of the problem state and problem rules, they are able to utilize the hint with some effectiveness, even when the solution times are very short to start with.

The hint study and Tower hint study, taken together, argue that the

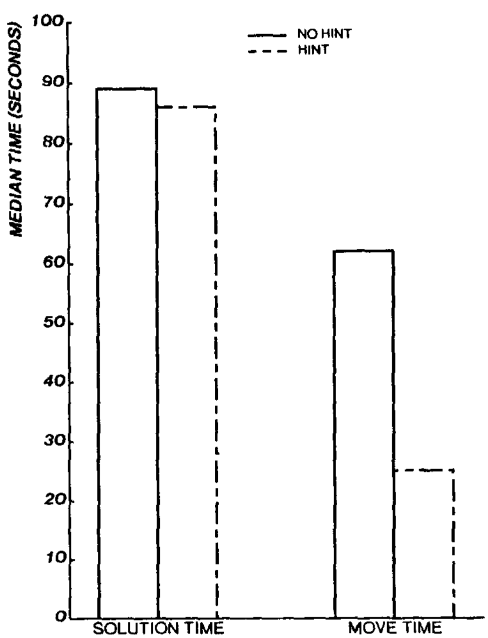

FIG. 11. Tower of Hanoi hints: Experiment 5B.

subjects' entry into the phase we have labeled the final path behavior is contingent on something other than their ascertaining the correct direction in which to proceed. The finding that they can utilize a hint if memory load is low (Tower problem) but not when memory load is high (Monster Move or Change problems) is evidence for the alternative explanation: that the subjects lack the ability to plan two moves. In order to test this assertion further, we performed an experiment to measure the problem difficulty for subjects who presumably already knew how to make moves. The hypothesis is that they should be able to solve the problem very rapidly, and that the Move problem and the Change problem should be approximately equivalent in difficulty, as they are in the final path portion of the normal problem-solving situation.

EXPERIMENT 6: RESTART STUDY

Method

Subjects. The subjects were 10 Carnegie–Mellon University students who received course credit for their participation.

Procedure. The subjects were given a Move (TA) or Change (CA′) problem (Fig. 1) with the usual written instructions. They were timed in solving the problem, and their verbalizations were recorded. As soon as they generated a solution, they were asked "to please try another problem." They were then handed a new sheet of instructions with another problem, giving the initial positions of the globes, and stating that "this problem uses the same rules for making moves as were used in the problem you just solved." The subject, after reading the instructions, began to solve the problem. Figure 12 shows the problem space for problem TA, and for the problem that followed TA, the "restart" problem. The pair of problems for CA′ had equivalent problem spaces. The follow-up problems were chosen so that there was no overlap between the correct or shortest solution path for the first and the second problems; that is, so that memory of solution path information would not help the subject, and could not explain any reduction in solution times.

Results

The results are presented in Table 13, which shows that the solution times for the original problems were close to the usual 2:1 Change:Move difficulty ratio. The actual ratio was 1.78:1, and a Bayesian analysis that compared the likelihood of the ratio being 2:1 with the likelihood of it being 1:1, even with a conservatively estimated prior probability of .8 for the 2:1 prior ratio, yielded a probability of $p < .05$ that the current probability was 2:1 rather than 1:1. The time to solve the follow-up problem was almost exactly the same for Move and Change problems, and was significantly less than the time required for the initial problem [$F(1,8) = 21.44$, $p < .0025$]. The number of moves was also significantly reduced in the follow-up condition [$F(1,9) = 5.45$, $p < .05$], being approximately equal for Move and Change problems, and less than half of what it was for the original problems. We conclude that the subject, knowing how to make moves from the prior problem, can transfer that knowledge and

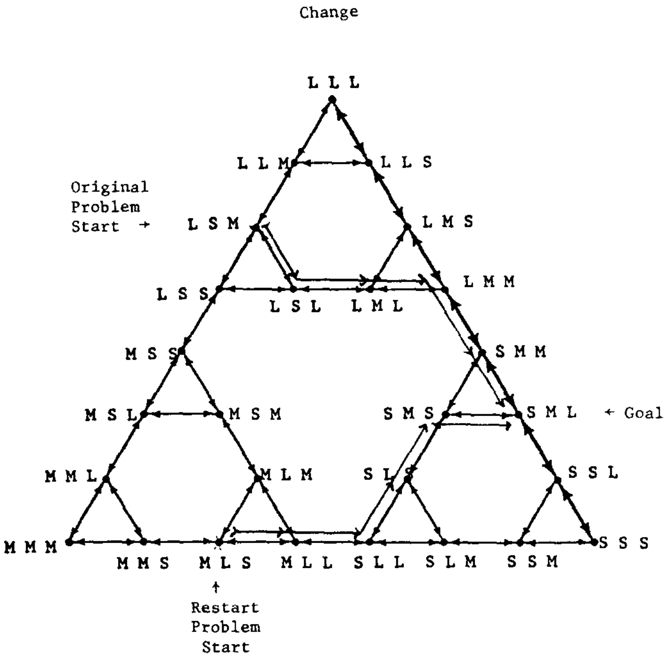

FIG. 12. Problem space for change problem space showing the nonoverlapping minimum solution paths for the original and restart problems.

thus accomplish the necessary planning, even though the solution path in the second problem is totally different. It appears that subjects without prior experience have to explore the problem space until they automatize move making so as to free up enough working memory to enable them to plan ahead.

DISCUSSION: MEMORY LOAD AND PLANNING BEHAVIOR: THE SOLUTION PROCESS

The conclusions that emerge when the studies reported above are considered as a whole are these:

- The problem rules (the restrictions on move legality) are a major source of problem difficulty and differences in difficulty (**Rule Learning Hypothesis**).
- Rule compatibility is a major influence on transfer of training (**Rule Similarity–Transfer Hypothesis**).
- Rules consistent with real world knowledge aid problem solving—as does the physical representation in the Tower problem—by reducing the memory load imposed by the rules (**Real World Knowledge Hypothesis**).

TABLE 13
Results of Restart Experiment

	Problem type and order			
	Move 1st	Move 2nd (restart)	Change 1st	Change 2nd (restart)
Total solution time				
Mean	905.4	346.6	1613.4	368.4
Median	761	351	1173	225.6
Move time				
Mean	725.4	290.8	1287	299.2
Median	659	277	933.5	242
No. moves				
Mean	29.6	12.4	24.8	13.2
Median	22	10	19	11
No. errors				
Mean	6	1.4	4.8	1.4
Median	4	1	5	2
Avg. time/move	24.5	23.5	51.9	22.7

- Providing a good external representation—the physical monsters—aids problem solving, and removing information from the external representation (the Dish-Move problem) hinders problem solving (**External Memory Hypothesis**).
- Intensive training on the rules, even in isolation from any problem setting or knowledge, significantly decreases the difficulty of the problems (**Rule Application Hypothesis**).
- Spatial or positional information is somehow privileged. It is easier to test objects at a single locus rather than at two loci, easier to keep track of changes in position than to track changes in nonspatial attributes of an entity, and easier to plan subgoal–goal pairs of moves involving only one rather than two *imaged* entities (**Spatial Memory Load Hypothesis**).

In addition, it appears that subjects must go through a substantial amount of exploratory behavior before they reach a point where they can proceed quickly toward the goal. Their final behavior consists of a fairly standard sequence of moves organized in two bursts or chunks. The first-phase behavior of moves without progress can be viewed as a kind of rule familiarization process. It is posited that this process, by automating the rules, allows for planning to occur. Automating the rules removes some of the load of remembering the current state, the goal, the intervening state, and the legality tests involved in compiling two moves.

Our analysis receives substantial support from other experiments reported in the literature. We wish to discuss a few of these now, in their relation to our work.

Memory load and planning. The work of Klahr and Robinson (1981) with children further supports the memory load analysis presented here. They constructed a physical external representation of the Move problem that built into the physical structure Rule 3 as well as Rule 2. With both rule constraints built in, children were often able to plan ahead, often describing the entire five-move solution path that would be needed for a solution: something our adult subjects never did.

Memory for spatial information. The idea that spatial information is easier to remember than nonspatial is somewhat supported by studies on errors in iconic storage (Henderson, 1972; Von Wright, 1968, 1970) and on spatial chunking processes that operate when information is read out of the icon (Bartram, 1978). In addition, the outcome of a number of studies has suggested the existence of a separate or additional spatial working memory component which could account for the memory advantage of spatially organized information. (See, for example, Baddeley, Grant, Wight, & Thompson, 1974; Baddeley & Hitch, 1974; Brooks, 1967, 1968; Kroll, Parks, Parkinson, Bieber, & Johnson, 1970). In a number of experiments reported here, the results also support this view of a spatial or positional effect. Thus Rule 2 in Move problems has repeatedly been found to be more rapidly read (Experiment 2), more rapidly used to test the legality of moves (Experiment 2), and less error prone (Experiment 4A), than Move problem Rule 3, or Change problem Rule 2 or 3. The latter, unlike Move problem Rule 2, all involve tests at a distance rather than spatially contiguous ones.

Our experiments also indicate that working memory limits are less severe if memory contents are stable rather than changing. As we have pointed out, the Dish–Change problem, which was much easier than other Change problems, differed from them in one important way: while it did entail tests at a distance, it did not have an entity changing in size. A set of experiments related to this issue (Yntema, 1963; Yntema & Mueser, 1960, 1962) demonstrated that people are much more adept at keeping track of a number of abstract attributes of a single entity than they are at keeping track of the value of the same attribute of each of a number of entities. The authors attributed this result to the fact that in the harder task, the variables took on a state from the same set of possible states, thus requiring more retrieval cues and allowing interference between attributes. This result is analogous to our finding of Change problem difficulty, where the subject must keep track of the current state of three different entities (globes), with the variable (size) for each entity taking on one of three possible values. The monsters in the Move problem, on the other hand, take on states (globes) from a mutually exclusive set of states, thus reducing the interference.

In a recently reported series of experiments, Novick and Tversky (1983) studied the order in which subjects applied transformations in solving geometric analogies. In their problems, subjects had to apply a pair of transformations to a geometric figure in order to test the validity of the geometric analogy. They found a fairly consistent order of application in which location transformations were utilized first (corresponding to Moves in the current study), orientation transformations next, and finally transformations that introduced modifications (corresponding to Changes in the current study) were applied last. The order reported is, for the two relevant categories, the same as the difficulty order repeatedly found for Move and Change Monster problem isomorphs. The tentative explanation they offer for their results involves a hypothesized interaction of perceptual processing and working memory load: a suggestion close to the analysis offered for the results reported here.

A model of memory requirements. To examine further the soundness of our analysis for explaining the differences between the Move and Change isomorphs, we constructed a simple model of the two-move sequence for each problem type. The premise of the model is that while the problems are structurally isomorphic, the difference in the spatial and imaginal qualities of the problem representations might operate differently in Move and Change problems. (We have already indicated (Figs. 4 and 8) the maximum number of imaged entities that must be simultaneously envisaged in using the problem rules and in making subgoal moves, and the relationship of that number to Rule and subgoal difficulty.) The model incorporates an analysis of the total information necessary at each stage of planning and executing a two-move sequence and delineates perceptual and working memory components of that information. The stages include (1) attempting the goal move, (2) testing the legality of the goal move, (3) adopting a subgoal to clear a block of the goal move, (4) testing the legality of the subgoal move, (5) imaging the subgoal move, (6) testing the legality of the goal move in the situation generated by the imaged subgoal move, and finally (7) executing the subgoal and goal moves.

The model generally follows the perceptual strategy for the Tower of Hanoi outlined by Simon (1975), and has some relationship to the look-ahead portion of the model proposed by Karat (1982). The Simon and Karat models differ, however, in that they describe behavior in the Tower of Hanoi problem, which we have shown (in its three-disk version) to be much easier than the Monster problems. The focus of those models is thus on problems using more than three disks and thus involving larger goal stacks and much more elaborate subgoal information and execution. As we have shown, there are vast differences (on the order of 16 to 1) in difficulty between the Tower problem and the more difficult of the isomorphs we have focused on. The sources of difficulty are quite different.

The difference between the two problem types is perhaps best illustrated by Karat's analysis (for the Tower problem) of the look-ahead production in his model. That production, which plans and executes a three-move sequence to clear a pair of disks blocking a desired subgoal move, is described as performing a *trivial task*, and the moves are made relatively quickly. In contrast, our subgoal clearing move (Fig. 7) involves only a single move to clear one globe blocking a desired move, is achieved late in the solution process when subjects are in the final path portion of their behavior and even then is often error prone (Table 12). Egan and Greeno (1974) present an analysis of subjects' errors versus depth of goal stack that also shows the substantial planning capabilities evidenced by subjects in the Tower–Disk–Peg problem as compared with the minimal capabilities demonstrated in the more difficult of the isomorphs we have investigated.

With the help of our model, analysis of the two-move sequence for each problem type shows that the sequence for Change problems entails approximately 1.5 times as many steps as the same sequence for Move problems (due primarily to the necessity of performing more size comparisons in Change problems). What is more important, the memory load imposed by the Change problems exceeds that imposed by the Move problems, primarily because of comparison of objects at a distance. The imposed memory load is shown in Fig. 13 (aggregated over the seven stages of a two-move sequence for the two isomorphs). It can be observed that the Change problem imposes a memory load of six chunks 3 times as often as does the Move problem, a result fully consistent with the analysis developed here.

CONCLUSION

The work we have presented in this paper is focused on problem behavior that serves as a kind of front end to the move-making behavior reported by Karat (1982), Simon (1975), and Egan and Greeno (1974). We have examined the processing load engendered by different problem isomorphs that all eventually yield the same type of "expert" final path behavior in identical problem spaces. We have focused on the seemingly unimportant detailed features of the problem that impose processing and memory load, thus producing the vast range of difficulties of the various isomorphs.

We have shown that the solution process depends on becoming expert at utilizing the problem rules to make moves, that the automation of the

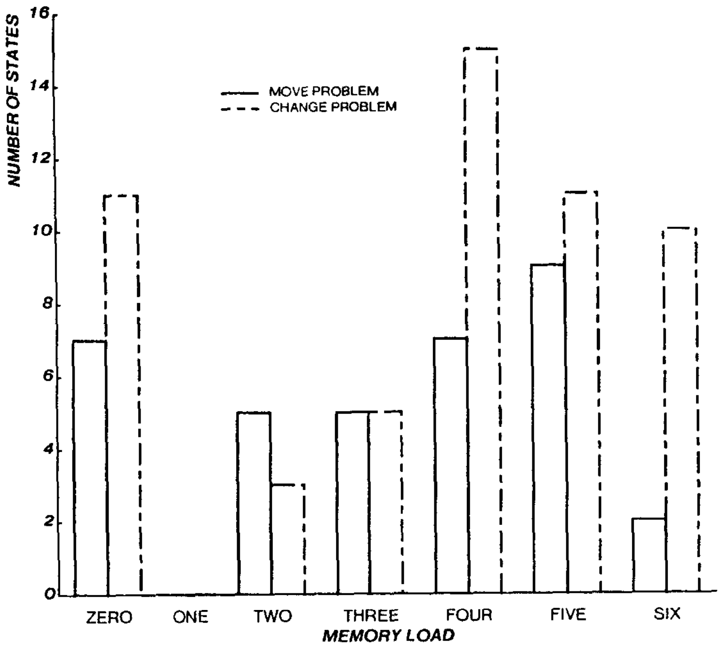

FIG. 13. Subgoal move model: the aggregate memory load imposed during the performance of a subgoal–goal move pair.

rule using behavior is a necessary precursor to planning behavior, and that a small amount of planning capability enables a rapid solution. We have further analyzed the various isomorphs in some detail so as to be able to understand and predict the processing load and thus the problem difficulty as measured by problem solution time. Problems that impose higher processing loads require more training for move automation to allow planning to occur, and thus require longer solution times.

In addition, we have moved between the macro- level of perceptual information and isomorph type to a microanalysis of the detailed move-making behavior of the subjects. This microanalysis of the final path behavior explains the temporal patterns of subjects' moves and the final path behavior sequence. It also incorporates an analysis of the spatial/imaginal comparisons that partly account for the difference between Move and Change isomorphs.

This analysis, while consisting of a detailed examination of a set of isomorphic problems, sets forth a view of the initial stages of the problem-solving process that may be applicable to a much broader array of problem types and also to a variety of pedagogical situations. The problem solver has a sharply limited amount of processing capacity that is easily overloaded by the memory requirements of unfamiliar problem rules. This overload prevents even minimal amounts of necessary planning from occurring. In problems where training and practice are possible, this limitation can be overcome through the automation of some of the knowledge. Repetitive drill on a delimited class of problems in physics or mathematics would be an example of a pedagogical strategy for overcoming this limitation. Allowing practice on a set of repeated problems that control the introduction of new information or strategies, would produce enough automation to allow planning, learning, and generalization to occur. On problems that do not allow for rehearsal, exploration of the problem space, or rule practice of some other type, difficulties similar to those encountered with different isomorphs would remain.

Parallel mechanisms are exemplified by the work of Wason and Shapiro (1971), Wason (1972), and Johnson-Laird et al. (1972) on logic problems where errors are sharply reduced by the introduction of problem isomorphs that involve familiar materials instead of unfamiliar materials (Wason & Johnson-Laird, 1972). The effect of increasingly complex rules in increasing error rates in a logic selection task has also been demonstrated in an experiment by Johnson-Laird and Wason (1970). Subjects' insight into the task (i.e., which cards would have to be selected to test an assertion) was not often all or none in nature, but rather, depended on the complexity of the rule being tested. That work, as well as the extensive series of experiments reported here, demonstrates the limitation of a problem analysis based on problem space structure alone, and argues for the necessity of a more complete analysis that incorporates features of problem space and move operator complexity with a description of the information processor and his/her limitations, in the context of a situation in which significant amounts of learning may alter the parameters as behavior proceeds.

REFERENCES

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, **86**, 124–140.

Baddeley, A. D., Grant, S., Wight, E., & Thompson, N. (1974). Imagery and visual working memory. In P. M. Rabbitt & S. Dornic (Eds.), *Attention and performance*. New York: Academic Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.

Bartram, D. J. (1978). Post-iconic visual storage: Chunking in the reproduction of briefly displayed visual patterns. *Cognitive Psychology*, **10**, 324–355.

Brooks, L. R. (1967). The suppression of visualization in reading. *Quarterly Journal of Experimental Psychology*, **19**, 289–299.

Brooks, L. R. (1968). Spatial and verbal components in the act of recall. *Canadian Journal of Psychology*, **22**, 349–368.

Carroll, J. M., Thomas, J. C., & Malhotra, A. (1980). Presentation and representation in design problem-solving. *British Journal of Psychology*, **71**, 143–153.

Chase, W. G., & Clark, H. H. (1972). Mental operations in the comparison of sentences and pictures. In L. W. Gregg (Ed.), *Cognition in language and memory*. New York: Wiley.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, **4**, 55–81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121–152.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, **3**, 472–517.

Cohen, N., & Corkin, S. (1981, October). *The amnesiac patient H.M.: Learning and retention of a cognitive skill.* Paper presented at the 11th Annual Meeting of the Society for Neuroscience, Los Angeles.

Egan, D. E., & Greeno, J. G. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving. In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, NJ: Erlbaum.

Ernst, G. W., & Newell, A. (1969). *GPS: A case study in generality and problem solving*. New York: Academic Press.

Gagné, R. M., & Smith, E. C. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, **63**, 2–18.

Hayes, J. R., & Simon, H. A. (1974). Understanding written problem instructions. In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, NJ: Erlbaum.

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory*. Erlbaum.

Henderson, L. (1972). Visual and verbal codes: Spatial information survives the icon. *Quarterly Journal of Experimental Psychology*, **24**, 439–447.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, **63**, 395–400.

Johnson-Laird, P. N., & Wason, P. C. (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology*, **22**, 49–61.

Just, M. H., & Carpenter, P. A. (1976). Verbal comprehension in instructional situations. In D. Klahr (Ed.), *Cognition and Instruction*. Hillsdale, NJ: Erlbaum.

Karat, J. (1982). A model of problem solving with incomplete constraint knowledge. *Cognitive Psychology*, **14**, 538–559.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem solving and planning processes on preschool children. *Cognitive Psychology*, **13**, 113–148.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. Short-term memory while shadowing: Recall of visually and aurally presented letters. *Journal of Experimental Psychology*, **85**, 220–224.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science (Washington, D.C.)* **208**, 1335–1342.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice–Hall.

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

Novick, L., & Tversky, B. (1983, November). *Transforming our knowledge of geometric analogies*. Paper presented at the 24th Annual Meeting of the Psychonomic Society, San Diego.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, **7**, 268–288.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, **8**, 165–190.

Von Wright, J. M. (1968). Selection in visual immediate memory. *Quarterly Journal of Experimental Psychology*, **20**, 62–68.

Von Wright, J. M. (1970). On selection in visual immediate memory. *Acta Psychologica*, **33**, 280–292.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard Univ. Press.

Wason, P. C., & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, **31**, 591–597.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, **23**, 63–71.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, **8**, 7–17.

Yntema, D. B., & Mueser, G. E. (1960). Remembering the present states of a number of variables. *Journal of Experimental Psychology*, **60**, 18–22.

Yntema, D. B., & Mueser, G. E. (1962). Keeping track of variables that have few or many states. *Journal of Experimental Psychology*, **63**, 391–395.

(Accepted March 11, 1985)