Bulletin of the Psychonomic Society
1984, 22 (4), 301-304

Subgoal length versus full solution length in predicting Tower of Hanoi problem-solving performance

HERMAN H. SPITZ, SHULA K. MINSKY, and CANDACE L. BESSELLIEU
E. R. Johnstone Training and Research Center, Bordentown, New Jersey

To determine whether performance on three-disk Tower of Hanoi problems is related to solution length (number of steps to transfer all disks to their goal pegs in a minimum path sequence) or to length of the first subgoal (number of steps to transfer the first disk to its goal peg in a minimum path sequence), we devised two sets of problems in which solution length was held constant but subgoal length was varied by one step. The performance of groups of nonretarded children and retarded young adults on these problems indicated that subgoal length was the primary variable. Unexpectedly, initial response latency (planning time) was unrelated to performance level.

Within the framework of information-processing theory, the Tower of Hanoi (TOH) problem is frequently cited as representative of transformation or search problems (Greeno, 1978; Simon, 1983). The material for the TOH can vary considerably, but typically consists of a number of differently sized disks that are distributed on three pegs in a given (start-state) configuration. The problem solver must rearrange the disks by moving them one at a time on the pegs, never placing a larger disk on a smaller one, in order to reach a goal-state configuration. Each TOH transformation can be accomplished in a minimum number of steps, which constitutes optimal solution.

In order to perform efficiently, subjects must make a means-ends analysis *prior* to giving their first response (Newell & Simon, 1972). Consequently, it is fair to say that the capacity to search ahead mentally a number of moves is crucial to performance on transformation problems such as the TOH. We refer to this capacity as the subject's "depth of search." In the present context, depth of search is the number of moves that a subject can mentally search, up to but not including the final move to the desired subgoal or goal (since the final move of simply transferring the freed disk to its goal peg is presumed to be automatic).

Everyone's depth-of-search capacity has some limit, as does any component of working memory. When the solution to a transformation problem requires a large number of steps, so that it is difficult or impossible for the problem solver to search the entire sequence from the start to the goal state, a frequently efficient strategy is to establish a subgoal, or series of subgoals, on the path to solution (Wickelgren, 1974). Recent studies suggest that many children from 6 to 10 years old and many mildly retarded adolescents and young adults attempt to solve three-disk TOH problems by first reaching an initial subgoal, such as transferring the largest disk to its correct goal peg (Borys, Spitz, & Dorans, 1982; Klahr & Robinson, 1981; Spitz, Webster, & Borys, 1982).

The available evidence, then, suggests that it is the number of steps to the subgoal, not the number of steps to the final goal, that determines the difficulty of the TOH problem at this level of performance. But the evidence is inferential; what is required is a more direct test, which can be made by setting up problems in which the number of steps to solution is constant while the number of steps to the first subgoal varies by one step. Half of all possible six-step problems require four steps to transfer the first disk to its goal peg in the correct sequence, the remaining half require only three steps, and most do not have alternative paths to optimum solution. Similarly, at reduced step levels, varying-length subgoals are available for four- and five-step problems, creating the opportunity to use four- to six-step problems for a direct test of the hypothesis that it is subgoal length, not solution length, that determines the difficulty of TOH problems for the maturationally young.

Finally, there is evidence that when college students perform verbal problem-solving tasks, the better performers spend relatively more time in planning their strategy than do the poorer performing students (Sternberg, 1981). Because our TOH apparatus is interfaced with a computer, we have the opportunity to determine whether planning time is related to TOH performance and to intelligence when young children and retarded young adults are the problem solvers.

This research was supported in part by NICHD Grant HD 15069. Suzanne Borys was instrumental in the design of the experiment. We thank Joseph Miller, principal, and the students of the Roebling Elementary School, as well as the students of the E. R. Johnstone Training and Research Center, for their generous help. Send reprint requests to the first author, E. R. Johnstone Training and Research Center, Bordentown, NJ 08505.

METHOD

Subjects

There were two groups of 20 subjects. The first group was composed of children from the regular third and fourth grades of a neighboring public school. Eight boys and 12 girls [mean chronological age (CA) = 9.3 years, SD = 0.6 years] were drawn from the list of volunteers and were transported two at a time to the research quarters of the Johnstone Center, where they were tested individually. The second group was composed of residents of the Johnstone Center, a relatively short-term state facility for mildly and moderately retarded adolescents and young adults. They were included in order to determine whether results with nonretarded children had sufficient generality to be generated also by a group approximately equal in mental age (MA), but quite disparate in CA and intelligence. Additionally, their performance would provide information on whether planning-time and pause-time patterns during problem solving are related to intelligence. Eight young men and 12 young women agreed to participate (mean age = 18.9 years, SD = 3.09 years; mean IQ = 60, SD = 5 years; mean MA = 9.58 years, SD = 1 year).

Apparatus and Procedure

The Tower of Hanoi apparatus was interfaced with a Byt-8 microcomputer in an adjoining room. The Johnstone TOH apparatus consists of three silver metal rods, 9 mm in diameter, that extend 193 mm from a black wooden base that is 7 cm high, 16 cm deep, and 25 cm wide. The diameter of the large disk is 76 mm, that of the medium disk is 52 mm, and that of the small disk is 35 mm, with interior holes 12 mm in diameter. The experimenter and the subject have essentially identical materials, but only the subject's TOH apparatus is interfaced with the computer.

Small magnets, embedded inside each of the three disks, activate magnetic reed switches embedded inside the pegs. When a disk is moved off or onto a peg, it initiates a series of switch closures that are processed by the computer, which registers the start state, the latency (in .01 sec) from when the problem was uncovered to the inital move, the latencies between moves, the peg from which a particular disk is moved and the peg to which it goes, the move time, and the total time elapsed after each move. When the problem is completed, the program computes the average pause time, average move time, total time, and total number of moves. This information is displayed on the CRT screen and simultaneously stored on a disk for later printout.

Prior to presenting each problem, the experimenter placed a cardboard screen between the subject and the apparatus so that she could set up both the start state (on the subject's board) and the goal state (on the experimenter's board). When the subject was alerted, the experimenter lifted the cardboard screen and simultaneously pressed the Start button on a control box. The computer then assumed all timing and recording functions until the experimenter pressed the Stop button when the subject had solved the problem, violated a rule, or given up. During testing, at least one observer was in the adjacent room to monitor the recording equipment and to observe (through a one-way mirror) and record the behavior and comments of the subjects.

Instructions for the two-disk criterion task. Prior to being presented the experimental problems, the subjects were required to solve relatively simple two-disk TOH problems. First, they had to demonstrate that they knew which was the large and which was the small disk and that they understood the rules that had been explained to them. The experimenter then set up a two-disk, three-step problem requiring transfer of the disks from the leftmost to the rightmost peg. To reach criterion and advance to the three-disk problems, the subjects had to solve both this problem and its inverse (from the rightmost to the leftmost peg) twice consecutively. If a subject continually failed these problems, the experimenter demonstrated a problem in which two disks on the center peg were transferred to an end peg, following which the subject started over on the initial problem. Although it took some time for some of the retarded subjects to pass criterion (primarily because they violated the rules), all eventually passed, as did all the nonretarded children. However, three retarded subjects were replaced because of bizarre behavior, a procedural mistake, and computer failure.

Instructions for the three-disk problem. The medium-sized disk was added to the other two, and the subjects had to demonstrate that they knew which was the largest, which the smallest, and which the medium-sized disk, and understood the rules, which were repeated to them. The experimenter then replaced the screen and set up the first three-disk problem. Prior to each trial, the subject was told how many moves were required to solve the problem. Subjects who solved the problem in excess moves and were to receive the same problem again were told how many moves they had made, and encouraged to try to do it in fewer moves.

Two sets of problems were given. The three problems in each set are designated 1, 2, and 3 in Table 1. Note that all six problems had the same goal state, and that the three problems in each set required, respectively, four, five, and six steps to solution. What differentiated the sets was the start-state configurations and, most importantly, the number of steps to the first subgoals. The Shorter Subgoal set required one, two, and three steps to the first subgoal, whereas the Longer Subgoal set required two, three, and four steps to the first subgoal. Half the subjects in each group received the Shorter Subgoal set first and then the Longer Subgoal set; the other half received the opposite order. Each problem was given three times consecutively, so that the subjects received a total of 18 trials. One point was scored for each successful solution, defined as reaching the goal state in the minimum number of moves. When a subject violated a rule, the trial was stopped, the violation was pointed out, and that trial was scored as zero.

Table 1
Start and Goal State Configurations

Order	Start State			Goal State			No. Steps to First Subgoal	No. Steps to Solution
				Shorter Subgoal				
1	_	L	S M	M L	_	S	1	4
2	_	S L	M	M L	_	S	2	5
3	M	S L	_	M L	_	S	3	6
				Longer Subgoal				
1	S	M	L	M L	_	S	2	4
2	S	_	M L	M L	_	S	3	5
3	_	_	S M L	M L	_	S	4	6

Note—L = large disk, M = medium disk, and S = small disk.

RESULTS

Performance

The performance results are given in Figure 1, in which it is immediately evident that the performance of both groups was primarily related not to the number of steps to solution, but to the number of steps to the first subgoal. These results were statistically reliable. A 2 (groups) x 2 (order) x 6 (problems) mixed ANOVA produced reliable effects for groups [$F(1,36) = 22.95$, $p < .001$] and for problems [$F(5,180) = 15.23$, $p < .0001$]. There were no other reliable effects and no reliable interactions. The group of children had a mean score of 7.2 on the Shorter Subgoal set, reliably higher than the mean score of 5.4 on the Longer Subgoal set [$t(19) = 5.38$, $p < .001$]. Scores for the retarded adolescents were 5.0 and 3.4, respectively [$t(19) = 3.06$, $p < .01$]. For these subjects, then, the problems with shorter subgoals were easier to solve than the problems with subgoals that were one step longer, despite the fact that the number of steps to solution was equated.

Pause times. The median pauses for the first successful trial of each of the six problems are given in Figure 2 for the nonretarded group and in Figure 3 for the retarded group. All curves are similar in showing a 4- to 6-sec pause prior to the first move, with the remaining moves following rapidly.

Consequently, only this initial pause prior to making the first move (referred to as planning time) was examined in depth. Planning time did not differentiate successful from unsuccessful trials. Each subject's initial pauses were averaged over all his or her first successful trials on each of the six problems (or on those of the six problems on which at least one success was obtained). For the retarded and nonretarded groups, the medians of these averaged pauses were 5.7 and 5.4 sec, respectively.

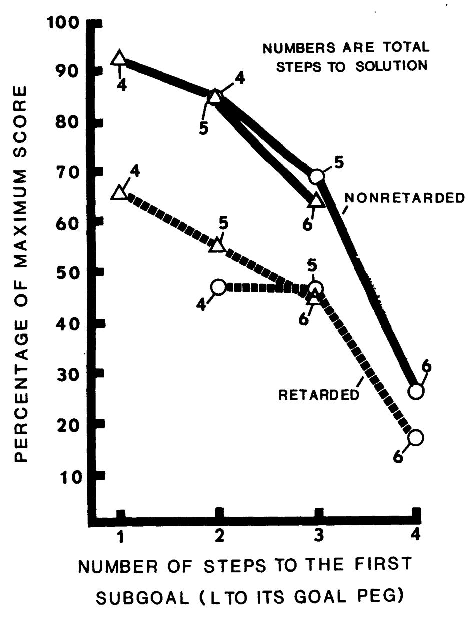

Figure 1. TOH performance of retarded and nonretarded groups as a function of number of steps to the first subgoal (given on the abscissa) versus number of steps to solution (given at the data points). Triangles indicate Shorter Subgoal set; circles indicate Longer Subgoal set.

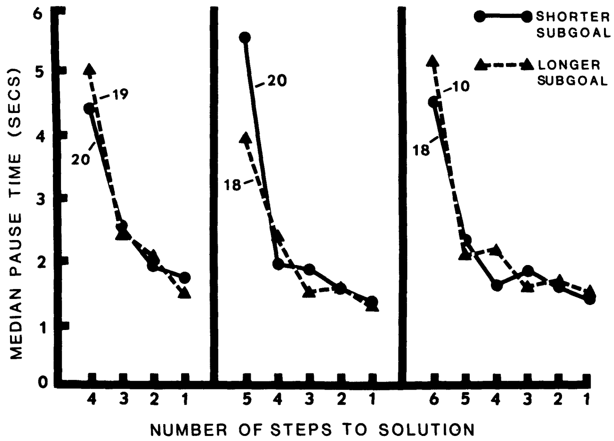

Figure 2. Median pause times of nonretarded subjects on first successful trial of each set of four- to six-step TOH problems. Number of subjects who generated each curve is given in the figure.

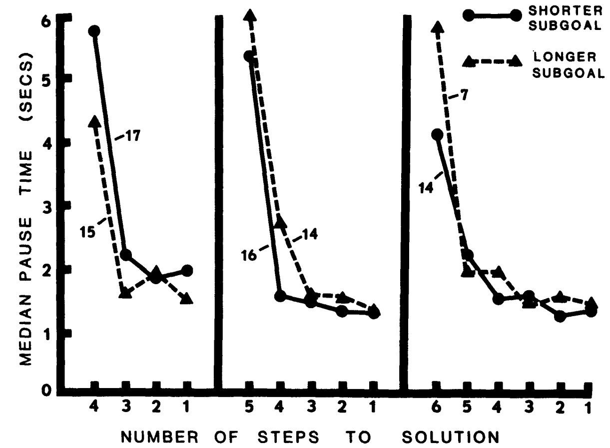

Figure 3. Median pause times of retarded subjects on first successful trial of each set of four- and six-step TOH problems. Number of subjects who generated each curve is given in the figure.

The same data for the first unsuccessful trials were 5.6 and 4.3 sec, respectively. None of these differences was reliable, either between (using a rank test) or within (using a signed-rank test) subject groups.

We averaged each subject's initial pauses over all 18 trials; the median planning time of the retarded group was 5 sec, and that of the nonretarded group was 4.5 sec, a difference that was not reliable. Each subject's initial pauses on the first trial of each problem, irrespective of outcome, were averaged; for the retarded group, the median was 5.4 sec, and for the nonretarded group, it was 6.1 sec, a difference that was also not statistically significant.

The median for the average planning times of the 20 retarded subjects on the Shorter Subgoal set was 5.2 sec and on the Longer Subgoal set was 5.5 sec. For the nonretarded group, the respective medians were 4.8 and 5.8 sec. Only the difference for the nonretarded group was reliable [$T(10,10) = 51.5$, $p < .05$], indicating that they took somewhat longer before making their first moves on the problems having longer initial subgoals than on problems having initial subgoals that were one step shorter.

Because the two sets of problems were counterbalanced, it was possible to determine the effect of subgoal length when subjects were first confronted with a three-disk problem. The median planning times on the first trial for the four-step problem requiring two steps to the first subgoal were 9.9 and 10.8 sec for 10 retarded and 10 nonretarded subjects, respectively. These contrast with median planning times that were about half as long, 4.8 and 4.9 sec, for the remaining subjects in each group whose first problem required only one step to the first subgoal. The differences were reliable for the nonretarded group [$T(10,10) = 73$, $p < .05$] and marginally reliable for the retarded group [$T(10,10) = 81$, $p < .10$]. When these same problems were the fourth problems in a set, the first-trial planning times did not differ reliably, being about 4 sec in both conditions for the nonretarded group and 4 to 6 sec in both conditions for the retarded group.

Correlational analyses corroborated the lack of any important relationship between planning time and performance. We calculated 24 correlations of performance scores and planning times, for the two subject groups separately and combined, on Shorter and Longer Subgoal sets and for successful and unsuccessful trials. None of the correlations was statistically significant. We also correlated the performance scores of each subject with his or her average planning time on all 18 trials. The correlations were unreliable for the groups analyzed separately, but for the combined 40 subjects the correlation was a significant $-.33$ ($p < .05$). This modest negative correlation indicates that the more successful subjects tended to take somewhat *less* time before making their initial moves than did the less successful subjects.

DISCUSSION

When persons have had limited experience with a problem, and when the number of steps from start to goal state is beyond their search capacity, it is reasonable to expect them to break the problem into subgoals, and the most likely major subgoal in the TOH problem is the transfer of a single disk to its goal peg (perhaps with smaller subgoals nested within the major one). For this subgoal to be reached, the target disk must be freed to move and the target peg must be cleared of any obstructing disks. The execution of these moves requires a means-ends analysis that is at the heart of TOH problems. For tower-to-tower configurations, Simon (1975) described this means-ends execution in some detail, and Karat (1982), using only tower-to-tower configurations, proposed a model in which the first subgoal is the transfer of the largest disk to its goal peg, with subsequent subgoals being the transfer of the largest disk not yet on the goal peg. (Of course, this strategy is not applicable when some disk other than the largest disk must reach its goal peg first, e.g., when the start state is M L S and the goal state is L S M.)

One might expect that those subjects who took a longer time to plan their moves would make fewer errors than subjects who responded more quickly, but this was not the case either for the comparisons of individuals within each group or for the comparison between groups. Only when both groups were combined did a modest, but negative, correlation emerge, indicating that subjects who took less planning time performed somewhat better than subjects who took more planning time. Although the nonretarded children increased their planning time by a median of 1 sec on the Longer Subgoal problems compared with the Shorter Subgoal set, this did not compensate for problem difficulty and the children nevertheless performed more poorly on the Longer Subgoal set. One second of additional planning time did not compensate for the need to search an additional step.

REFERENCES

Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology*, **33**, 87-110.

Greeno, J. G. (1978). Nature of problem-solving abilities. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 1. Human information processing* (pp. 239-270). 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 ability in preschool children. *Cognitive Psychology*, **13**, 113-148.

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

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

Simon, H. A. (1983). Search and reasoning in problem solving. *Artificial Intelligence*, **21**, 7-29.

Spitz, H. H., Webster, N. A., & Borys, S. V. (1982). Further studies of the Tower of Hanoi problem-solving performance of retarded young adults and nonretarded children. *Developmental Psychology*, **18**, 922-930.

Sternberg, R. J. (1981). Intelligence and nonentrenchment. *Journal of Educational Psychology*, **73**, 1-16.

Wickelgren, W. (1974). *How to solve problems: Elements of a theory of problems and problem solving*. San Francisco: Freeman.

(Manuscript received for publication March 9, 1984.)