Retention of Problem Solutions: The Re-solution Effect

Roger L. Dominowski

University of Illinois at Chicago

Linda S. Buyer

Governors State University

Running head: THE RE-SOLUTION EFFECT

This research was supported in part by a grant from the University of Illinois Research Board. Appreciation is expressed to Anders Ericsson and an anonymous reviewer for their comments on an earlier version of this paper. Correspondence should be addressed to Roger L. Dominowski, Psychology Department (m/c 285), University of Illinois at Chicago, 1007 W. Harrison St., Chicago, IL 60607 (e-mail: rdomin@uic.edu).

ABSTRACT

Four experiments compared the re-solution performance of prior solvers versus prior nonsolvers given the correct solutions. The first two experiments challenged Weisberg and Alba's (1981) contention that solving a problem and being shown the solution yield equivalent problem knowledge. In both experiments, students who initially had solved problems showed near-perfect recall of the solutions after a one-week delay, far superior to recall by students who had been shown the correct answers. In experiment three, solvers showed poor solution retention when the connection between the problem and the solution was not meaningful. Experiment four showed that, with meaningful problems, solvers and those merely provided with solutions have qualitatively different problem representations. The findings can be explained in terms of differential understanding of problems and their solutions.

Retention of Problem Solutions: The Re-solution Effect

The bulk of problem-solving research has focused on the initial discovery of solutions, with relatively little attention paid to how well solutions are remembered. The scarcity of research on solution memory is surprising because the topic involves important theoretical issues. Historically, a contrast existed between associative and Gestalt approaches to problem solving. Associative theorists (e.g., Thorndike, 1911) viewed problem solving as the acquisition and elicitation of associative bonds. From a Gestalt perspective (e.g., Kohler, 1925/1976; Maier, 1940; Wertheimer, 1945/1982), it was accepted that problems could be solved by direct application of past experience, but emphasis was given to another form of problem solving that involves the reorganization or restructuring of experience and knowledge (see Ohlsson, 1984).

Insight refers to the new knowledge of problem structure that follows reorganization. It was held that solutions involving restructuring and insight would be remembered well (see Osgood, 1953; Woodworth & Schlosberg, 1954). Indeed, in Kohler's studies with apes, quite skillful solutions were frequently observed when a problem was presented a second time, which implies excellent solution memory. In contrast, in Thorndike's puzzle-box experiments with cats, many trials were required before efficient solutions occurred, which implies that little was learned or remembered from solving a problem once, and which led to the view of problem solving as a gradual, trial-and-error acquisition process. Although both associative and Gestalt approaches were later applied to human problem solving, researchers did not explore the differential expectations regarding solution memory. In a general discussion of human and animal problem solving, Scheerer (1963) repeated the Gestalt position, asserting that solutions accompanied by insight will be easily retained.

In a complete rejection of Gestalt views of problem solving, Weisberg and Alba (1981) specifically questioned the validity of Scheerer's assertion, stating that he presented no supporting data. More important, Weisberg and Alba reported that, after a one-week interval, only 65% of their subjects solved the nine-dot problem when it was presented a second time. They concluded that their data contradicted the Gestalt prediction and that solution memory showed quite ordinary forgetting, consistent with their overall position that problem solving involves primarily retrieval of information from long-term memory, with no role for insight. The adequacy of Weisberg and Alba's retention finding may be questioned. Their subjects, rather than having solved the problem on its first presentation, had failed to solve and been given the solution. The solution had been acquired, not via problem solving but in a manner similar to a standard memory task. Consequently, the appropriateness of their data for characterizing retention following successful problem solving is at best unclear (Dominowski, 1981).

Selected aspects of work on the generation effect are potentially relevant. The basic generation effect is illustrated by Jacoby's (1978) study. He gave subjects simple associative puzzles (e.g., given foot s _ _ e, complete the second word) and found that recall of the solution words was better when subjects had generated them, compared to when subjects had read the solution words (e.g., shoe) immediately prior to puzzle presentation. Jacoby's results show that generating an answer leads to better recall than reading an answer, but his results may not be directly relevant to Weisberg and Alba's (1981) finding. Weisberg and Alba's subjects had tried unsuccessfully to solve the nine-dot problem before receiving the solution, a condition that Jacoby did not explore. This condition was studied by Slamecka and Fevreiski (1983), using materials similar to Jacoby's (e.g., find the opposite of pursue, given av _ _ d). They found that, compared to a condition in which answers were read, recall of solution words was better when subjects attempted to generate answers, whether or not they were successful. For present purposes, the most important finding was that there was no recall difference following successful versus unsuccessful generation attempts. That finding supports Weisberg and Alba's contention that whether a subject solves, or fails to solve but is given the solution, is of no consequence. Several accounts of the generation effect, although differing about details, agree that the effect depends on the item being represented in memory as a familiar unit (e.g., Slamecka & Fevreiski, 1983; Gardner & Hampton, 1985). As such, they are compatible with the view that problem solving is nothing more than cued retrieval of information from memory (e.g., Weisberg & Alba, 1981). The view outlined above stands in clear contrast to that of the Gestalt psychologists. Their emphasis on restructuring differentiates problem solving from ordinary memory. Wertheimer (1945/1982) commented on the difference between solving a problem via restructuring and being shown a solution after failing to solve. He downgraded the latter as unlikely to result in re-creating the process of restructuring necessary for real understanding. Recently, Metcalfe (1986a, 1986b) found differences between problem solving and memory, with her results generally supporting a restructuring process in problem solving. From this perspective, one would expect solvers to differ from nonsolvers who are given solutions.

Buyer and Dominowski (1989) found a retention difference favoring solvers of number-phrase riddles. In that experiment, subjects had to complete phrases referring to "facts" about numbers, e.g., "64 = Squares on a Checkerboard" from partial cues (e.g., "64 = S on a C"). Successful generation led to far better retention than unsuccessful generation (with solution feedback), which was no better than mere reading of the items. Although subjects had to already know the relevant number facts in order to complete the phrases, it is clear that the phrases as items of general knowledge are more complex, less familiar, and less well-integrated than the words used by Slamecka and Fevreiski (1983). The solutions to the nine-dot problem and other problems are still more complex, do not exist in subjects' semantic memories prior to problem presentation, and gain meaning only in the context of the problem situation. A plausible inference is that, for such problems, solvers' retention will be decidedly superior to that of nonsolvers who are given the solution. The present studies provide the data necessary to evaluate this expectation.

To summarize the issues: From a memory-retrieval viewpoint, performance when problems are presented a second time should be the same for prior solvers as it is for nonsolvers who have been given solutions, and both groups should show less-than-perfect performance. From a restructuring viewpoint, solvers should be "near-perfect" and better than nonsolvers who have received solutions. To make the data of this study comparable to those of Weisberg and Alba (1981), a one-week retention interval was used. To provide some generality to the results, several problems were employed, including the nine-dot problem used by Weisberg & Alba.

EXPERIMENT 1

Method

Subjects and Design

The subjects were 50 students from introductory psychology at the University of Illinois at Chicago who participated in partial fulfillment of a course requirement. Each subject participated in two sessions that were separated by a one-week interval; subjects attempted seven problems in Session 1, and were given the same seven problems in Session 2. Session 1 performance was used to identify solvers and nonsolvers, separately for each problem; comparison of Session 2 (retention) performance by the two groups constituted the primary contrast.

Materials

Seven problems were selected from published reports; they were chosen to provide some variety, to be amenable to paper-and-pencil presentation, and to yield a reasonable number of initial solutions within a few minutes' working time. The problems were: The Prisoner, Horse Trade, Nine Dot, Train, and Farmer problems (Maier & Casselman, 1970), the Card problem (Davis, 1964), and the "gold-dust" version of a water-jars problem (Restle & Davis, 1962). Each problem was presented on a single sheet of paper that included space for recording solutions. Three of the problems required drawing line patterns to meet problem constraints (prisoner, nine-dot, and farmer problems). The horse-trade and train problems require interpreting a paragraph of information to arrive at a correct numerical answer (amount of money made, number of trains seen on a trip); to be counted as correct, subjects had to provide an adequate explanation of how they arrived at their numerical answers. The card problem requires interpretation of a series of sentences (e.g., "To the right of a king there is a queen or two.") to determine the identities of three playing cards aligned in a row. The water-jars problem requires finding a series of "transfers" between jars having stated capacities to arrive at a designated, measured amount.

Procedure

Subjects signed up for two sessions. At the beginning of Session 1, they were told that the research project required giving each person a number of problems, some of which they would receive in this session, and some during a second session a week later. No mention was made that the Session 1 problems would be repeated in Session 2.

Session 1 data were collected using small groups. For each problem, work sheets were distributed face down (these sheets contained complete instructions and work space); a start signal was given, and subjects were given 3 - 5 minutes to reach a solution (working time varied across problems). When time was up, a stop signal was given, work sheets were collected, and then the solution to the problem was shown via overhead projector. Solution feedback remained in view as long as at least one subject indicated a desire to study it further. This procedure was followed for each of the seven problems.

Subjects were seen individually for their second sessions. They were given the same seven problems, using the same problem order and time limits as in Session 1. Solution feedback was given if subjects failed to solve a problem. After the seventh problem, subjects were told the purpose of the study and were asked to assist further with the research by refraining from discussing the experiment with other students.

Results and Discussion

The prisoner problem and the train problem yielded one and two initial solutions, respectively; these problems will not be considered further, and the results concern performance on the other five problems. The data of primary interest concern performance by prior solvers versus prior nonsolvers when problems were presented a second time after a one-week delay.Solver and nonsolver groups were defined separately for each problem. Figure 1 shows the overall performance levels in Session 2 for the two groups. Panel (A) indicates that prior solvers showed near-perfect retention, re-solving problems at a 98% rate, which was much higher than the percentage of solutions by prior nonsolvers (61%), $z = 6.27$, p

<ㅃ.001. Across problems, prior solvers were uniformly successful in reproducing solutions, whereas success rates for prior nonsolvers varied. The solver-nonsolver difference in solution rate was statistically significant for each of the horse trade, card, water jars, and nine-dot problems ($p < .05$). The one exception was the farmer problem, which was solved in Session 2 by 95% of prior nonsolvers. It should be noted that for the nine-dot problem used by Weisberg and Alba (1981), Session 2 performance by prior solvers (100%) was far superior to that by prior nonsolvers (38%), $z = 2.66$, $p < .01$.

Insert Fig. 1 about here

As shown in Figure 1(B), when prior nonsolvers did succeed in Session 2, they were slower than prior solvers in producing solutions, t (48) = 3.63, $p < .001$.[1] Figure 2 shows mean solution times for the two groups for each of the five problems. Prior solvers were faster for every problem, and solution times varied across problems; the solution-time difference between groups was statistically significant for all but the nine-dot problem. These solution times, because they concern only solutions actually produced, do not completely represent performance differences between groups across problems. If minimum solution times (time limit + 1 sec) were added for all those failing to solve in Session 2, the gap between prior solvers and nonsolvers would be substantially increased. For the farmer problem, which nearly all subjects in both groups solved in Session 2, the data in Figure 2 show that prior solvers (mean = 13.8 sec) were decidedly faster than prior nonsolvers (mean = 32.1 sec): t (46) = 2.54, $p < .02$. Application of a two-factor analysis of variance to the solution time data gave no suggestion of a problem by group interaction.

Success or failure on any particular problem was independent of performance on other problems. Phi coefficients varied from -0.14 to +0.43 with a median value of +0.06. Therefore, solver and nonsolver groups did not differ systematically in any general problem-solving abilities.

Insert Figure 2 about here

The results are clear with respect to the main issues: in terms of solution probability, prior solvers showed virtually no forgetting over a one-week period. In addition, prior solvers were faster than prior nonsolvers who had been given the solution. The findings support Wertheimer's proposal that reorganization of problem information leads to excellent memory for the solutions. They indicate that Weisberg and Alba (1981) overestimated the amount of solution forgetting by people who have solved a problem and contradict their proposal that generating a solution and being given a solution are equivalent. But they do agree with Buyer and Dominowski's (1989) findings that self-generation produces superior recall. The implications of the results will be discussed after all the experiments are described.

Experiment 2 was a constructive replication of the first experiment, addressing some procedural issues that concerned primarily the poor Session 2 performance by prior nonsolvers. Because Session 1 data were collected in group sessions, it was possible that some nonsolvers paid inadequate attention to the solution feedback. In addition, for two of the problems, the feedback might have been technically incomplete. For both the prisoner and nine-dot problems, the completed solution consists of a pattern of lines; subjects had been shown the completed pattern but not how it is drawn, so it might be argued that they had not been given all of the solution. Regarding prior solvers, the first experiment did not, strictly speaking, provide data showing that they had learned and remembered anything from their initial solutions to the problems. Although this notion seems unlikely to be correct, Session 1 solution times are needed for comparison with Session 2 re-solution times; a reduction in solution times for prior solvers would unambiguously demonstrate that they had learned and remembered useful information from their first solutions. In the second experiment, therefore, individual first and second sessions were used, with a comparison of two types of feedback given to nonsolvers of the prisoner and nine-dot problems in Session 1.

EXPERIMENT 2

Method

Subjects and Design

Fifty University of Illinois at Chicago (UIC) students from introductory psychology participated in partial fulfillment of a course requirement. Each subject participated in two individual sessions held one week apart, attempting seven problems in Session 1 and the same problems in Session 2. For two problems (prisoner and nine-dot), two forms of solution feedback (static vs. dynamic) were used in Session 1, with approximately half of the non-solving subjects receiving each form.

Materials and Procedure

The seven problems and the procedures used in Experiment 1 were again employed, with the following changes. Individual first sessions were conducted, with solution times recorded for solvers and with nonsolvers receiving solution feedback on a sheet of paper to be examined for as long as they desired. For the prisoner and nine-dot problems, a schedule was created to randomly assign subjects to receive either static or dynamic solution feedback in Session 1. Static feedback consisted of the pre-drawn solution (as in Experiment 1), whereas for dynamic feedback, the experimenter drew the solution lines in full view of the subject (and repeated doing so if requested). Because some subjects solved these problems and thus did not require feedback, each form of feedback was received by approximately half of the nonsolving subjects. In Session 2, half the subjects were asked if they had worked on any of the problems during the week between sessions and, if so, to describe what they had done.

Results and Discussion

As in Experiment 1, the data are based on five problems (excluding the prisoner and train problems which had a total of three initial solutions) with solvers and nonsolvers defined separately for each problem. The overall results replicated the findings of Experiment 1. As shown in Figure 3(A), prior solvers were 99%

successful in re-solving problems presented in Session 2, compared to only 55% solutions for prior nonsolvers, $z = 6.77$, $p < .001$. The difference between groups was statistically significant for the card, water jars, and nine-dot problems ($p < .05$, one-tailed). For the farmer and horse-trade problems, prior solvers were 100% successful, but prior nonsolvers were only slightly worse in Session 2 (96% solutions for the farmer problem, 94% for the horse-trade problem). As in Experiment 1, Session 2 performance by prior solvers of the nine-dot problem (100%) was much higher than that for prior nonsolvers (46%), $z = 2.66$, $p < .01$.

Replicating a finding in Exp. 1, success on any one problem was independent of success on other problems, with phi coefficients ranging from +0.01 to +0.28 with a median value of +0.05. This finding indicates that solver and nonsolver groups did not differ in general problem-solving abilities. Rather, the act of solving a problem has important consequences for later performance.

Insert Figure 3 about here

Figure 3(B) shows that, when Session 2 solutions were achieved, prior nonsolvers were generally slower than prior solvers in producing them, $t(48) = 3.05$, $p < .01$. Figure 4 presents solvers' and nonsolvers' mean solution times for the individual problems; as in Experiment 1, solvers were faster on all problems. The solution-time difference between groups reached statistical significance levels for the card and farmer problems. These data reflect only solutions actually produced; inclusion of minimum solution times for those who failed to solve would yield huge differences between prior solvers and nonsolvers for the card, water jar, and nine-dot problems. For the farmer problem, which was solved in Session 2 by nearly all prior nonsolvers, solvers (mean = 17.5 sec) were nonetheless markedly faster than nonsolvers (mean = 53.3 sec) in producing solutions: $t(47) = 3.29$, $p < .01$. Application of a two-factor analysis of variance did not suggest a problem by group interaction.

One purpose of Experiment 2 was to determine whether solvers learned anything from their initial solutions. On the average, solvers' Session 2 solution times were more than 80 sec. faster than their first solutions. Re-solution times were faster in 96% of the cases, confirming the expectation that solvers would acquire useful information from their initial solutions. Across problems, there was some variation in the size of the reduction (52-139 sec.), which reflected differences in initial solution times, i.e., problems with longer initial solution times tended to have larger time reductions in Session 2. Within problems, first solution times were poor predictors of second solution times; correlations ranged from -0.17 (card) to +0.44 (horse trade) with a median value of +0.22.

Two problems had sufficient numbers of initial solvers to allow a comparison of fast versus slow session-one solvers (using median splits) with respect to changes in solution times between sessions. For the Horse problem, fast and slow solvers' (n = 9 each) initial solution times were vastly different (38.3 and 121.1 sec, respectively), but their re-solution times were systematically shorter and differed only slightly (19.6 and 36.1 sec, respectively). For the Farmer problem (n = 18 per subgroup), fast (68.6 sec) and slow (143.6 sec) solvers' initial times were quite different, whereas their resolution times were much shorter and more similar (11.5 and 23.6 sec, respectively). These data suggest that reaching a solution is more important than how long one took to find the solution, with respect to later performance on a problem. Experiment 4 allowed us to examine this phenomenon in greater depth because we collected data regarding the students' problem representations. Thus, we were able to explore both the relation of solving/not solving in Session 1 to the characteristics of the problem representations and the relation of the representations' characteristics to second session performance.

Insert Figure 4 about here

The type of solution feedback given to nonsolvers of the prisoner and nine-dot problems in Session 1 had no effect on Session 2 performance. Static vs. dynamic feedback led to Session 2 solution rates of 42% and 46% for the prisoner problem, 50% and 43% for the nine-dot problem, respectively. Thus, allowing nonsolvers to observe the drawing of solution lines did not improve Session 2 performance. Also, providing solution feedback on an individual basis in Experiment 2 did not benefit nonsolvers, compared to Experiment 1 where group feedback was used. In Experiment 2, the nonsolvers' Session 2 performance was slightly poorer and the solver-nonsolver difference was slightly larger than in Experiment 1.

Subjects reported thinking about or attempting to work on a small minority of problems (15%) during the one-week interval between sessions. No subject reported directing any attention to a solved problem during the interval. In Session 2, the solution rate for "worked-on" problems (43%) was equivalent to the overall solution rate for prior nonsolvers (48%). Efforts to work on unsolved problems during the interval were generally not beneficial; less than one third of the reported attempts were described as successful (subjects sometimes could not reproduce enough of the details of a problem to make a meaningful attempt). When attempts were described as successful, the problem was always solved in Session 2. Consequently, a minority (about 15%) of the Session 2 solutions produced by prior nonsolvers actually represented re-solution of a problem solved during the interval between sessions. The results replicate the findings of Experiment 1 in regard to the near-perfect retention shown by prior solvers, the large differences between prior solvers and prior nonsolvers, and differences across problems. The data also showed that re-solutions by prior solvers are substantially faster than their original solutions, indicating considerable learning and retention. Varying solution feedback procedures did not lead to improved subsequent performance by nonsolvers, suggesting that neither inattention to feedback nor static solution feedback is responsible for failures by prior nonsolvers when problems are presented again.

Experiment 3 was designed to extend these findings by exploring the consequences of having individuals solve problems in which the relation between the problem and the solution was arbitrary. We hypothesized that, if Wertheimer (1945/1982) was correct in asserting that restructuring plays a critical role in the retention and application of problem solutions then it should follow that there will be no memorial benefits derived from correctly solving problems that have arbitrary solutions. On the other hand, if Weisberg and Alba (1981) are correct in their assertion that restructuring is irrelevant to the retention of problem solutions, then we should find that memory for previously solved problems is near-perfect (as we found in Experiments 1 and 2) even when no restructuring is possible.

EXPERIMENT 3

Contrary to the predictions of Weisberg and Alba (1981), the data from Experiments 1 and 2 provide support for the Gestalt position regarding the role of restructuring as a causal factor in solution retention and ease of application (Kohler, 1925/1976; Maier, 1940; Wertheimer, 1945/1982). Experiment 3 was intended to extend these findings by showing that the "re-solution effect" disappears when there is no meaningful relation between the problem and its solution. Thus, for this experiment the problems and their solutions were paired randomly.

Method

Subjects

Thirty students from the University of Illinois at Chicago participated in the experiment for credit in an introductory psychology class. They were told that the experiment was about extra-sensory perception (ESP).

Materials

Forty-eight randomly selected, concrete words (taken from Paivio, Smythe & Yuille, 1968) were divided randomly into two sets of 24 words each. One of the sets was designated as the "cue words." The other was designated as the "response words." Then, one response word was randomly designated as the "correct" associate for each of the cue words. The 48 words were presented to subjects on a single sheet of paper with the 24 cue words arrayed in a column on the left side of the page. There was a blank (___) next to each cue word. The column was divided into eight groups of three words each. For each group of three cue words, the three correct response words were randomly arrayed in a row on the right side of the page. This arrangement was used so that subjects would average about one-third correct responses.

Procedure

Subjects were told that word pairs had been formed by randomly pairing words together. The subjects were then instructed to use their ESP to select the correct associate from among the response terms for each set of three cue words (and write it in the blank next to the cue word). Subjects were next shown a list of the correct word pairs and asked to estimate the percentage of people who would guess those two words would be paired. The subjects were asked to estimate the percentages to make sure that they actually processed each of the correct word pairings (Stage 1). Subjects returned one week later and were asked first to recall (Stage 2) and then to recognize (Stage 3) the correct response terms. In Stage 2, students were first presented with a sheet of paper containing the cue words in a column on the right side of the page and asked to write the correct response terms in blanks next to the cue words. The cue words were listed in the same order they had appeared in Session 1. Subjects were told to leave a blank empty only if they were absolutely certain that they could not recall the correct response word. In Stage 3, subjects were presented with a sheet of paper that differed from that used in Stage 2 by the inclusion of an alphabetical list of response words in a column on the left side of the page. Subjects were told to match the correct cue-response word pairs by using all of the response words only once and filling in all of the blanks.

Results and Discussion

We first checked to make certain that all subjects had correctly paired at least some of the cue-response word pairs presented in the initial session. In Stage 1, subjects correctly guessed approximately one-third of the correct response terms (mean = 29.9%, sd = 9.4%) and no subject guessed fewer than 12.5% of the correct responses. Thus, for each subject, we obtained second session data on both "solved" and "unsolved" problems.

Recall performance in the second session of this experiment was nothing like the recall performance in the second sessions of Experiments 1 and 2. The recall rates (10% for correctly guessed items, 3.5% for incorrectly guessed items) are substantially worse than the recall rates found in the first two experiments

(99% and 98% for correctly solved problems and 55% and 61% for incorrectly solved problems), although we again found that self-generated correct responses were better recalled than responses that had merely been provided ($t(29) = 2.49$, $p < .05$). In fact, in this experiment, the probability of guessing the correct response term in Stage 1 was greater than either the probability of correctly recalling a guessed item in Stage 2 ($t(29) = 6.1$, $p < .01$) or of recognizing it in Stage 3($t(29) = 1.973$, $p = .058$).

Stage 3 data again produced results similar to those found in the standard generation effect paradigm (e.g., Buyer & Dominowski, 1989; Slamecka & Fevreiski, 1983; Gardner & Hampton, 1985), i.e., 22% of the response terms which had been correctly identified in Stage 1 were recognized while only 11.9% of the response terms not correctly identified in Stage 1 were recognized ($t(29) = 3.51$, $p < .01$). However, the recognition rates from this experiment (22% and 11.9%) are extremely low when considered in the light of the recall rates from Experiments 1 and 2. This is especially true because, other things being equal, one usually expects better recognition than recall performance.

The hypothesis we derived from Gestalt theory was that if we eliminated meaningful relations between problems and their solutions we would dramatically reduce memory for solutions to correctly solved problems. The enormous difference in recall rates between the current experiment and the first two experiments suggests that when no restructuring is possible (e.g., when the relation between a problem and its solution is arbitrary) there is virtually no retention benefit to having previously solved a problem. Thus, these results again support the Gestalt position that restructuring plays an important role in retaining and applying problem solutions and serves to contradict Weisberg and Alba's (1981) contention that restructuring is an irrelevant component of solution memory.

EXPERIMENT 4

The data from our first three experiments can be summarized as follows: when a problem solution is not arbitrary, second session performance is virtually perfect following success in the first session and prior solvers outperform prior nonsolvers; when a problem solution is arbitrary (and there is nothing to understand about it, as in Experiment 3), second session performance is abysmal even by those who earlier correctly solved the problems. A remaining question involves the nature of the difference between those who have and those who have not successfully solved a meaningful problem. Experiment 4 was conducted to address this question.

In Experiment 4, both written and verbal protocols were collected so that we could explore any differences in problem representations between solvers and nonsolvers. Based on both Gestalt and Information Processing theories we expected to find that the representations of solvers were structurally different than those constructed by nonsolvers. Specifically, we expected to find that the problem representations constructed by solvers reflected the inherent structure of the problems, that they were well organized, that they accurately encoded the objectively defined problems, and that they were connected to related knowledge already present in the subjects' memories.

To examine a broad range of problems, problems representing Greeno's (1978; Greeno & Simon, 1988) problem typology were used. Greeno defines problem types according to the cognitive abilities required to solve the problems. The three basic problem types are: a) problems of inducing structure, e.g., analogies, which require identifying relations among components and fitting relations into patterns; b) transformation problems, e.g., Water Jars, which require planning skill based on a means-end strategy; and c) arrangement problems, e.g., Cryptarithmatic, which require composition skills and constructive search. In the current experiment, subjects were presented with one of each of Greeno's three problem types in Session 1. Subjects were asked to solve the identical problems again, in a second session, one week later.

Method

Subjects

Subjects were 54 undergraduates from the University of Illinois at Chicago who participated in the experiment for credit in an introductory psychology class.

Materials

Each of the three problem types Greeno (1978) identified was represented by two isomorphic exemplars of each type. (Problems that are isomorphic are structurally identical to one another.) Katona's (1940) card-trick problem was used as an arrangement exemplar, Luchin's (1942) water jars problem was used as a transformation exemplar, and inducing structure problems were represented by the Bourne, Dominowski, and Loftus (1979) "Matching" problem. A coin was flipped to assign each of the pairs of isomorphic problems to a "set" of problems (1 each arrangement, transformation, and induction) to be given to subjects to work on in the first session. In Session 1, approximately 50% of the subjects received one of the sets and the remainder received the other set.

Subjects were asked to think aloud while they were working. A tape recorder was used to collect the verbal protocols (concurrent verbal protocols = verbalizations while problem solving; retrospective protocols = post-problem verbalizations). The concurrent verbal protocols were collected for a purpose unrelated to the present experiments and will not be mentioned further. The verbal protocols from three subjects were lost due to mechanical failure of the tape recorder used to collect the data. Subjects were provided with the problems on separate sheets of paper and asked to show their work. Each sheet had room on it for the solution.

Procedure

Subjects in this experiment were run individually in both sessions. To acquaint subjects with the thinking-aloud procedure, each subject was first asked to count the number of windows in their house (Ericsson and Simon, 1993). No subjects expressed discomfort with this procedure. Any time a subject was silent for more than 10 seconds, he or she was told to "keep talking."

In Session 1, each subject was asked to solve one each of the induction, transformation, and arrangement problems. Each problem was presented with a time limit. A pilot study was performed to find time limits for each problem such that 50% of the subjects could solve the problem within the limit. Time limits ranged from 3.5 to 13 minutes. When subjects either solved each problem or the time allotted for the problem ran out, the problem sheet was removed, the subjects were shown the correct solution (if necessary), and subjects were asked the following questions about the problem.

1. In your own words, would you please describe this problem to me (that is, what are the instructions)?

2. Would you describe, in detail, what steps you took to solve the problem? (If the problem was not solved, the question was rephrased as."…爀what steps were you taking to try to solve the problem?").

3. Does this problem seem at all similar, in any way, to any other problem or situation you've ever encountered?

The answers to these questions constituted the retrospective verbal protocols. Two graduate students rated these protocols for (a) correspondence to the objectively defined problem (based on the answers to the first question), (b) coherence of the subject's internal representation (based on the second question), and (c) the number of connections between knowledge of the current problem and other knowledge (based on the third); and the same two students independently rated the written protocols (the subjects' worksheets) for the presence or absence of structural understanding (Wertheimer, 1945/1982). These measures are detailed in Appendix A.

Two subjects never wrote anything while working on any of the three problems. Four additional subjects failed to show their work for one of the three problems. If subjects failed to solve a problem within the time limits, they were allowed to view the correct answer until they indicated that they were satisfied.

Subjects were asked to return one week later and asked to solve the same problems, in the same order, again. In the first session the necessity for the second session was explained by saying that there were too many problems to do in a single session and that more problems (with no specification as to their nature) would be presented the next week.

Results and Discussion

In the initial session, the arrangement, transformation, and inducing structure problems were solved by, respectively, 32, 38, and 37 of the 54 subjects. As we found in Experiments 1 and 2, students were much more likely to solve a problem in the second session if they had successfully solved it in Session 1. Ninety-four, 97, and 100% of the prior solvers re-solved, respectively, the arrangement, transformation, and inducing structure problems. Only 67% (arrangement), 71% (transformation), and 69% (inducing structure) of those who initially failed to solve succeeded in Session 2. These percentages are approximately the same as the comparable initial solution percentages (59%, 70%, and 68.5%). Separate analyses of variance for each problem type using success in Session 1 as a predictor of solving in Session 2 were all significant (Smallest $\underline{F}(1, 52) = 7.58$, $\underline{MSe} = .13$, $\underline{p} = \leq .01$).

As in Experiments 1 and 2, solvers were faster at producing solutions in the second session than were initial non-solvers (mean difference for the arrangement problems = 79.37 sec., mean difference for the transformation problems = 30.3 sec., mean difference for the inducing structure problems = 29.5 sec.). Only the $\underline{t}$-test for the transformation problems reached significance ($\underline{t}(46)= 3.7$, $\underline{p} < .01$). However, the other $\underline{t}$-tests approached significance (arrangement problems: $\underline{t}(43)= 1.85$, $\underline{p} = .07$; inducing structure problems: ($\underline{t}(46)= 1.93$, $\underline{p} = .06$). Re-solutions in Session 2 were all substantially faster than the initial solution times in Session 1 (mean difference in first and second solution times for the arrangement problems = 225 sec., mean difference for the transformation problems = 67.2 sec., mean difference for the inducing structure problems = 48.5 sec.: smallest $\underline{t}$ (36) = 4.99, $\underline{p} < .01$).

Although solvers and nonsolvers in Experiments 1 and 2 didn't appear to differ in general problem-solving ability, different problems were used in Experiment 4, and I is possible that session 1 solvers are "good problem solvers" while session 1 non-solvers are "poor problem solvers." Total number of problems solved in Session 1 and speed of initial solutions both seemed to be reasonable indicators of general ability and were used as predictors. To adjust for differences in problem difficulty, solution speeds were indexed by the mean of the proportions of the maximum allowed time that was used to solve a problem. Because solving only one session 1 problem, however quickly, might reflect a different caliber of problem solver than does solving all three session 1 problems quickly, the interaction was included in the analysis as a third predictor. The criterion variables were solution speeds in session 2. Because this is a repeated-measures analysis, only individuals who solved all three problems in session 2 are included (n = 39). The univariate between-groups analyses indicated that none of the three measures of problem solving ability predicted speed of session 2 solutions (largest F (1, 35) < 1, all p's are n.s.). None of the multivariate, within-groups analyses revealed any interaction effect of the problem types with general ability on session 2 solution speed (smallest Wilk's lambda (2,34) = .9, all p's n.s.).

Analysis of Experiment 2 data indicated that Session 1 solution times were poor predictors of Session 2 solution times. Because the majority of the participants in Experiment 4 solved all of the problems in the second session (39 of 54), it was possible to relate initial solution speed for each problem to second session solution speed on all problems. Separately for each of the three problem types, Session 1 solution times were divided at the median into "faster" and "slower" groups. Thus, for each problem, we identified the solvers as faster or slower problem solvers. We then used the first session faster/slower designations of performance to predict speed of second session performance. None of the three repeated measures ANOVAs showed that being a faster or slower problem solver in the first session was predictive of second session solution times. Largest F (1,27) < 1; smallest Wilk's lambda's (2,26) = .815, all p's are n.s. Differences in session two solution rates (percentages, probabilities) and solution times between solvers and nonsolvers cannot be accounted for by differences in the solvers' general abilities.

Additional analyses were conducted to determine whether solving is associated with different problem representations than is nonsolving. Retrospective verbal protocols were analyzed because they best reflect the subjects' knowledge state after either solving or being shown a solution. Because the maximum values of the correspondence and coherence measures computed from the verbal protocols vary widely from problem to problem, they will be discussed in terms of percentages of the maximum for the remainder of this paper. Using solving in Session 1 as the predictor variable and the four measures of understanding (correspondence, coherence, connection to other knowledge, and structural understanding) as dependent measures, repeated measures analyses of variance were computed (separately for each of the three problem types) to determine whether the nature of their problem representations would serve to differentiate the solvers from the nonsolvers. All three repeated measures analyses showed that success in the first session was significantly related to the understanding measures (Smallest F (1, 48) = 4.85, MSe = .17, $p < .05$). The understanding measures for solvers and nonsolvers are shown in Table 1 below. Post-hoc single-degree of freedom contrasts showed that, for all three problem types, solvers were able to summarize the problems' solution more coherently than were the nonsolvers (Smallest F (1,48) = 5.38, MSe = .04, $p < .05$). In addition, the inducing structure and transformation solvers exhibited greater structural understanding than did their nonsolving counterparts (Smallest F (1,48) = 9.26, MSe = .19, $p < .01$) while the arrangement solvers were marginally better able to enumerate the problem's components (thus indicating greater correspondence) than were the nonsolvers (F(1,49) = 3.98, MSe = .03, p = .05). Thus, it is clear that solvers and nonsolvers differ primarily in terms of the degree of integration (as measured by structural understanding and coherence) of their stored representations.

Finally, an analysis of variance with the Session 1 Understanding Measures and Problem Type as predictors and the proportion of time required to solve a problem in session 2 as the dependent variable was computed. This analysis was conducted to determine whether the problem representations constructed by students as a result of their session 1 experiences would predict speed of second session solutions. The analysis indicated that all of the predictors, with the exception of the binary-coded structural understanding measure, account for variance in the second session solution times (correspondence: F(1,55) = 3.8, MSe = .05, p = .056; coherence: F(1,55) = 5.04, MSe = .05, $p < .05$; connection to other knowledge: F(1,55) = 3.46, MSe = .05, p = .068; problem type: F(1,55) = 5.56, MSe = .05, $p < .01$). Pearson correlations indicate that the relation of each of the understanding measures to session 2 solution times is in the expected direction. Greater correspondence (r = -.25, $p \leq .05$), more coherence(r = -.47, $p < .01$) and structural understanding (r = -.33, $p \leq .01$) are all associated with faster session two solution times.

Insert Table 1 about here

In summary, as in the first two experiments, those who succeeded in solving a problem in the initial session showed near-perfect retention of the solution after a one-week interval. Not only did prior solvers remember the solutions, they produced them much more quickly in Session 2 than they did in the initial session. Again, as in Experiments 1 and 2, they were also faster than those who had been shown the correct solutions after an unsuccessful solution attempt.

The analysis of the protocols indicated that solvers and nonsolvers have distinctly different problem representations after their initial experiences with the problems. The primary difference appears to be related to the coherence of the subjects' internal representations. This finding offers an explanation of the mechanism(s) accounting for the near-perfect solution retention for the prior solvers. It also provides strong support for Wertheimer's (1982/1945) contention that solving a problem produces an understanding that is different from that of someone merely shown the solution.

Further analysis of the protocols indicated that the quality of students' problem representations was predictive of second session solution times. Students were more efficient problem solvers in the second session if they had constructed representations in Session 1 that reflected the inherent structure of the problems, that were well-organized, that accurately encoded the objectively-defined problems, and that were connected to related knowledge already present in memory. Thus, solving in the first session is positively associated with the quality of the Session 1 understanding measures and these measures, in turn, are negatively associated with second session solution times. This finding provides further support for the contention that solving a problem and being shown a solution are qualitatively different experiences.

Recent work by Berardi-Coletta, Buyer, Dominowski, and Rellinger (1995) and Chi, de Leeuw, Chiu, and LaVancher (1994) has shown that eliciting self-explanations produces differences between those providing self-explanation and those that do not that mirror the observed difference between solvers and nonsolvers. Students that produce self-explanations produce better-integrated, more accurate mental models than control students. Thus, metacognitive prompting creates mental representations similar to those created by solvers. This suggests that the metacognitive activity implicit in self-explanation may cause the restructuring to which Wertheimer attributes the near-perfect recall.

GENERAL DISCUSSION

In Experiments 1, 2, and 4, solvers showed virtually perfect performance after a one-week retention interval and were decidedly superior to prior nonsolvers who had been given a solution. The findings are consistent with expectations based on Gestalt theory and contradict Weisberg and Alba's (1981) proposal that generating a solution and being given a solution should lead to equivalent subsequent performance. The contrast in retention rates between Experiments 1, 2, and 4 versus Experiment 3 supports Metcalfe's (1986a,b) distinction between problem solving and simple memory retrieval, and is compatible with the idea that at least some problem solving involves a restructuring process.

Slamecka and Fevreiski (1983) found no retention difference following successful versus unsuccessful attempts to generate target words. In contrast, we found large differences between solvers and nonsolvers in second session performance. This distinction between solvers and nonsolvers is, however, in complete agreement with the results of Buyer and Dominowski's (1989) study employing number-word phrases. Successful generation of the phrases led to far better retention than unsuccessful generation (with feedback), which was no better than mere reading of the items. Note that the problems and their solutions used by Buyer and Dominowski were more complex, less familiar, and less well integrated than the words used by Slamecka and Fevreiski. The problem solutions of the present study are still more complex, do not exist in subjects' semantic memories prior to problem presentation, and gain meaning only in the context of the problem situation. The pattern of results suggests that with solutions of even modest complexity, actual production of solutions will yield better subsequent performance than being given the solutions. Put differently, a positive effect from unsuccessful generation attempts might occur only with familiar words. It is worth noting that, in Experiments 1, 2, and 4, producing solutions led to much higher performance levels and a greater advantage than has typically been found in studies using word lists or other, similar verbal materials (as in Experiment 3). Indeed, questions have been raised concerning the robustness of the "generation effect" with word lists (e.g., Begg & Snyder, 1987;

Sutherland, Krug, Glover, 1988). Subjects in Experiments 1, 2, and 4 were not simply trying to remember words; rather, they were attempting to produce more elaborate behavior that would meet stated criteria within the constraints of a particular problem. As such, one might reasonably expect results different from those observed with simple word recall (as in Experiment 3).

Perfetto, Bransford, and Franks (1983) found that access to potentially useful information could be inhibited by prior learning of interfering material. Applying the interference notion in the present context, it is possible that prior nonsolvers performed poorly when given problems a second time because of interference from incorrect solutions they had produced during Session 1. The data do not support this idea, however. Examination of double failures, for Experiments 1 and 2, showed 9% repetitions of the same error, 49% different errors, and 43% cases to which the notion did not apply (e.g., no solution was offered on the initial failure). The improbability that interference was a major factor in prior nonsolvers' performance is further demonstrated by an examination of their performance on the horse-trade and train problems. For both problems, nonsolution virtually always involved offering an incorrect solution. Across the two experiments, only 37% of prior nonsolvers solved the train problem in Session 2, but 91% of prior nonsolvers solved the horse trade problem when it was repeated. This massive difference in performance cannot be explained in terms of interference, which was presumably maximal for both problems. Overall, the data imply that interference from earlier incorrect solutions was not an important influence on prior nonsolvers' Session 2 performance.

Metcalfe and Wiebe (1987) used patterns of "warmth ratings" to separate insight from noninsight problems. For insight problems, subjects' feelings of nearness to solution ("warmth") showed relatively abrupt transitions from low values to the maximum value at the point of solution. In contrast, warmth ratings exhibited a gradual increase over the course of work on noninsight or incremental problems. It was also found that subjects were less able to predict their performance on insight problems, compared to incremental problems. Both the difficulty of predicting success and the sharp changes in warmth are consistent with the idea that a process of restructuring takes place in solving insight problems. Three problems used in Experiments 1 and 2 were also used by Metcalfe and Wiebe (1987). The horse-trade and farmer problems were classed as insight problems, whereas the water-jars problem was categorized as an incremental problem. One aspect of the present findings showed marked differences among these problems. Re-solution times for prior solvers were the fastest for the horse-trade and farmer problems, whereas those for the water-jars problem were the slowest. The partitioning of these problems in terms of re-solution times replicated the categorization reported by Metcalfe and Wiebe.

Variation in re-solution times might reflect the kinds of differences in processing proposed by Metcalfe and Wiebe (1987). It would seem that re-reading the problem statement and literally executing the solution would require small and roughly constant amounts of time across problems. If so, then re-solution times would depend on the form of solution knowledge possessed by prior solvers. Greeno (1977, 1978) suggested that problems differ in the likelihood and form of understanding that might be acquired when solving them, and Sweller and Levine (1982) demonstrated that successful problem solvers can differ in the amount of knowledge about problem structure that they acquire. The data from our experiments suggest that, if the components of a solution can be coherently organized, as in the case of there being a key idea or insight involved, then re-solutions might be expected to occur rather quickly when the problem is presented again (as found in Experiments 1, 2, and 4). If on the other hand a problem lacks overall structure, if its solution consists of a number of weakly-integrated steps, then solvers' knowledge would be less organized and quick re-solutions would seem less likely (as found in Experiment 3).

In summary, our data indicate that there are meaningful relations between problem-solving processes and subsequent retention phenomena. The findings for nonsolvers (Experiments 1, 2, and 4) support the argument that providing solutions is not likely to yield understanding. Whereas solvers generate solutions from the features of the problem situation, relations between solutions and problem features may not be grasped when solutions are provided. Further, the data from Experiment 3 support the idea that meaningful problem-solution relations are very important for good solution retention. This interpretation of our findings provides a link between research on the generation effect and research into the consequences of self-explanation by providing a single explanation for all of the sets of findings.

Research on the generation effect has shown that the effect depends on the degree of self-generation (Buyer & Dominowski, 1989; Fiedler, Lachnit, Fay, & Krug, 1992). Research into self-explanations has shown that greater amounts of self-explanation produce better understanding of textual material (Chi, de Leeuw, Chiu, & LaVancher, 1994), that self-explanation produces more efficient problem learning and better ability to transfer solution knowledge (Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995; Buyer, Walsh, & Russell, 1997), and that it produces positive changes in problem conceptualization (de Grave, Boshuizen, & Schmidt, 1996). We believe that the positive effects associated with both self-generation and self-explanation result from the construction of complete, well-integrated, meaningful mental representations of the to-be-understood material. Whether externally provided solution feedback can be augmented to be equally effective is a matter for further research.

REFERENCES

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. Psychological Review, 94, 192-210.

Begg, I., & Snyder, A. (1987). The generation effect: Evidence for generalized inhibition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 553-563.

Berardi-Coletta, B., Buyer, L. S., Dominowski, R.L., and Rellinger, E. R. (1995). Metacognition and problem transfer: A process-oriented approach. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 205-223.

Bourne, L.E., Dominowski, R.L. and Loftus, E.F. (1979). Cognitive Processes. (Chapter entitled "Problem Solving"), Englewood Cliffs, New Jersey: Prentice-Hall.

Buyer, L. S., & Dominowski, R. L. (l989). Retention of solutions: It's better to give than to receive. American Journal of Psychology, 102, 353-363.

Buyer, L. S., Walsh, S., & Russell, T. A. (1997, May). Metacognitive training produces transfer across problem types. Poster presented at the 1997 American Psychological Society Convention in Washington, DC.

Chi, M.T.H., de Leeuw, N., Chiu, M.-H., LaVancher, C. (1994). Eliciting self-explanations improves understanding, Cognitive Science, 18, 439-477.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 669-675.

Davis, J.H. (1964). The solution of simple and compound word problems. Behavioral Science, 9, 359-370.

de Grave, W. S., Boshuizen, H. P. A., & Schmidt, H. G. (1996). Problem based learning: Cognitive and metacognitive problems during problem analysis. Instructional Science, 24, 321-341.

Dominowski, R.L. (1981). Comment on "An examination of the alleged role of 'fixation" in the solution of several 'insight' problems" by Weisberg and Alba. Journal of Experimental Psychology: General, 110, 193-198.

Ericsson, K.A., & Simon, H.A. (1993). Protocol analysis: Verbal reports as data (rev. ed.). Cambridge, Massachusetts: MIT Press.

Fiedler, K, Lachnit, H., Fay, D., & Krug, C. (1992). Mobilization of cognitive resources and the generation effect. Quarterly Journal of Experimental Psychology, 45A, 149-171.

Foos, P. W., Mora, J. J., & Tkacz, S. (1994). Student study techniques and the generation effect. Journal of Educational Psychology, 86, 567-576.

Gardner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 732-741.

Greeno, J.G. (1977). Process of understanding in problem solving. In N.J. Castellan, D.B. Pisoni, & G.R. Potts (Eds.), Cognitive Theory. Hillsdale, NJ: Erlbaum.

Greeno, J.G. (1978). Natures of problem-solving abilities. In W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes. Hillsdale, NJ: Erlbaum.

Greeno, J.G. & Simon, H.A. (1988). Problem solving and reasoning. In R. C. Atkinson, R. J. Hernstein, G. Lindzey, & R. D. Luce (Eds.), Stevens' handbook of experimental psychology. New York: Wiley.

Grosofsky, A., Payne, D. G., & Campbell, K. D. (1994). Does the selective generation effect depend upon selective displaced rehearsal? American Journal of Psychology, 107, 53-68.

Jacoby, L.L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. Journal of Verbal Learning and Verbal Behavior, 17, 649-667.

Johnsey, A., Morrison, G. R., & Ross, S. M. (1992). Using elaboration strategies training in computer-based instruction to promote generative learning. Contemporary Educational Psychology, 17, 125-135.

Katona, G. (1940). Organizing and Memorizing. New York: Columbia University Press.

Kohler, W. (1947). Gestalt Psychology. New York: Liveright.

Kohler, W. (1976). The Mentality of Apes. New York: Liveright. (Originally published 1925).

Luchins, A.S. (1942). Mechanization in problem solving: The effect of einstellung. Psychological Monographs, Whole No. 248.

Maier, N. R. F. (1940). The behavior mechanisms concerned with problem solving. Psychological Review, 47, 43-53.

Maier, N.R.F., & Casselman, G.C. (1970). Locating the difficulty in insight problems: Individual and sex differences. Psychological Reports, 26, 103-117.

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 288-294.

Metcalfe, J. (1986b). Premonitions of insight predict impending error. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 623-634.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. Memory & Cognition, 15, 238-246.

Myers, J. L. (1966). Fundamentals of experimental design. Boston: Allyn & Bacon.

Ohlsson, S. (1984). Restructuring revisited: I. Summary and critique of the Gestalt theory of problem solving. Scandinavian Journal of Psychology, 25, 65-78.

Osgood, C. E. (1953). Method and Theory in Experimental Psychology. NY: Oxford University Press.

Paivio, A., Smythe, P. E., & Yuille, J.C. (1968). Imagery versus meaningfulness of nouns in paired-associate learning. Canadian Journal of Psychology, 22, 427-441.

Perfetto, G.A., Bransford, J.D., & Franks, J.J. (1983). Constraints on access in a problem solving context. Memory and Cognition, 11, 24-31.

Restle, F., & Davis, J.H. (1962). Success and speed of problem solving by individuals and groups. Psychological Review, 69, 520-536.

Scheerer, M. (1963). Problem-solving. Scientific American, 208, 118-128.

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. Journal of Verbal Learning and Verbal Behavior, 22, 153-163.

Sutherland, J. A., Krug, D., & Glover, J. A. (1988). The selective displaced rehearsal hypothesis and failure to obtain the generation effect. Bulletin of the Psychonomic Society, 26, 413-415.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-end analysis and learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 463-474.

Thorndike, E.L. (1911). Animal Intelligence. New York: MacMillan.

Weisberg, R.W., & Alba, J.W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. Journal of Experimental Psychology: General, 110, 169-192.

Wertheimer, M. (1982). Productive Thinking. Chicago: University of Chicago Press. (Originally published 1945)

Woodworth, R. S., & Schlosberg, H. (1954). Experimental Psychology. (rev. ed.). NY: Holt & Co.

FOOTNOTES

---

[1]For the overall comparison of prior solvers versus nonsolvers, the mean solution time for the appropriate problem was subtracted from each individual solution time, producing adjusted times that were independent of problem differences. The statistical test was performed on the adjusted times, with degrees of freedom conservatively based on the number of subjects rather than the number of entries (Myers, 1966).

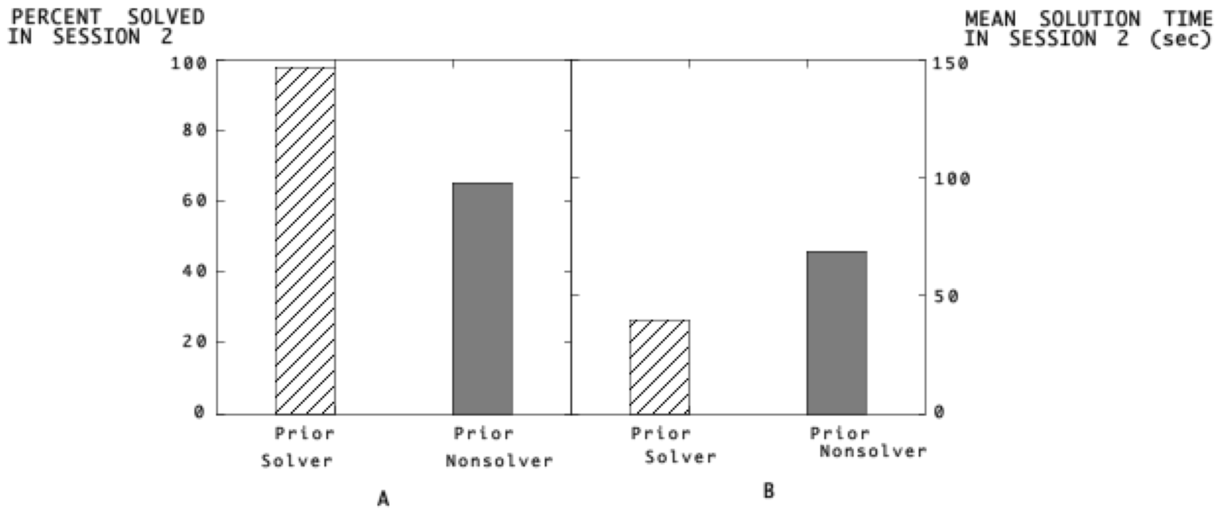
FIGURE CAPTIONS

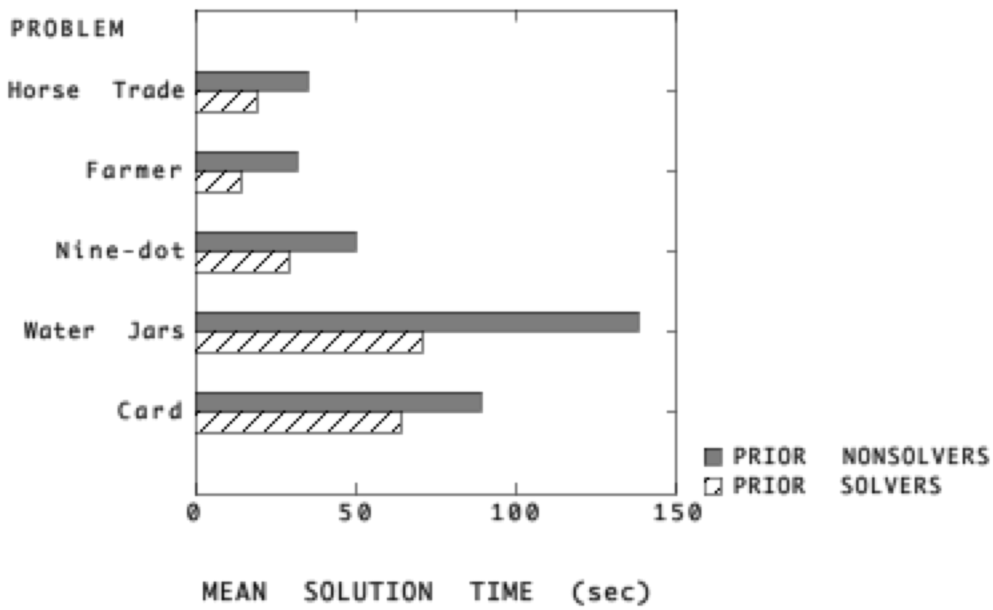
Figure 1. Performance of prior solvers and prior nonsolvers in Session 2, Experiment 1. Panel (A) shows the percentages of problems solved; panel (B) shows the mean solution times, based only on problems actually solved.

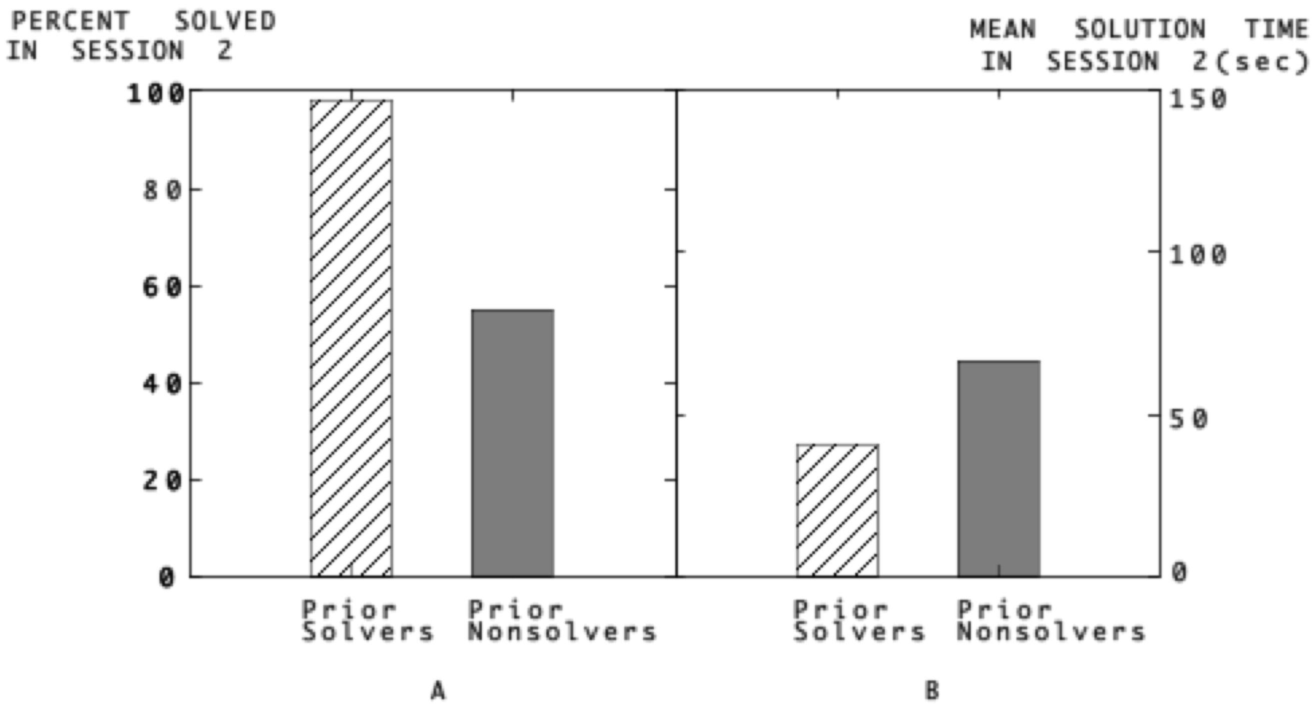
Figure 2. Second-session mean solution times for individual problems, Experiment 1; based only on problems actually solved.

Figure 3. Performance of prior solvers and prior nonsolvers in Session 2, Experiment 2. Panel (A) shows the percentages of problems solved; panel (B) shows the mean solution times, based only on problems actually solved.

Figure 4. Second-session mean solution times for individual problems, Experiment 2; based only on problems actually solved.


PERCENT SOLVED
IN SESSION 2
MEAN SOLUTION TIME
IN SESSION 2 (sec)
100
80
60
40
20
0
150
100
50
0
Prior
Solver
Prior
Nonsolver
Prior
Solver
Prior
Nonsolver
A
B


PROBLEM
Horse Trade
Farmer
Nine-dot
Water Jars
Card
PRIOR NONSOLVERS
PRIOR SOLVERS
0
50
100
150
MEAN SOLUTION TIME (sec)


PERCENT SOLVED
IN SESSION 2
100
80
60
40
20
0
Prior
Solvers
Prior
Nonsolvers
A
MEAN SOLUTION TIME
IN SESSION 2(sec)
150
100
50
0
Prior
Solvers
Prior
Nonsolvers
B

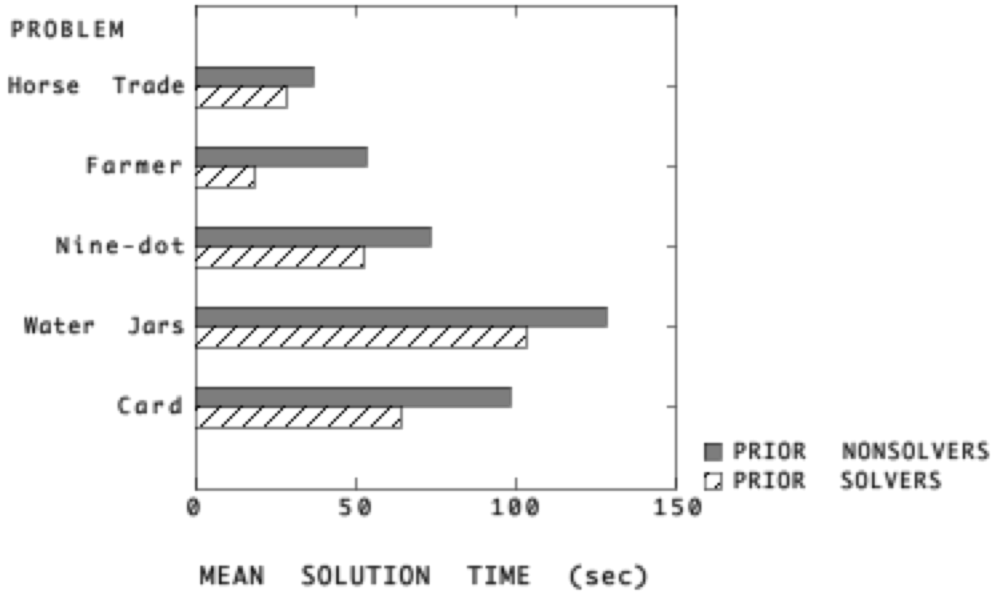

PROBLEM
Horse Trade
Farmer
Nine-dot
Water Jars
Card
PRIOR NONSOLVERS
PRIOR SOLVERS
0
50
100
150
MEAN SOLUTION TIME (sec)

Table 1

UNDERSTANDING MEASURES DERIVED FROM THE WRITTEN AND VERBAL PROTOCOLS

Understanding Measures	Status in First Session: Solver	Status in First Session: Nonsolver
Arrangement Problems		
	mean (sd)	mean (sd)
Corrected* Correspondence Rating	.96 (.07)	.87 (.26)
Corrected* Coherence Rating	.52 (.38)	.18 (.23)
Connection to Other Knowledge	.34 (.47)	.29 (.45)
Structural Understanding	.56 (.47)	.42 (.48)
Transformation Problems		
Corrected* Correspondence Rating	.87 (.16)	.93 (.10)
Corrected* Coherence Rating	.80 (.31)	.15 (.31)
Connection to Other Knowledge	.74 (.59)	.67 (.49)
Structural Understanding	.88 (.28)	.17 (.36)
Inducing Structure Problems		
Corrected* Correspondence Rating	.59 (.22)	.53 (.25)
Corrected* Coherence Rating	.30 (.21)	.16 (.15)
Connection to Other Knowledge	.71 (.52)	.75 (.48)
Structural Understanding	.53 (.48)	.13 (.34)

* Corrected means that the raw rating for each problem was divided by the maximum rating possible for that problem, i.e., it is a proportion of the total possible.

APPENDIX A
Protocol Analyses

Measure of Correspondence

Correspondence between the subjects' representations and the problems was coded by counting the number of basic problem facts included in their retrospective descriptions of each problem. Subjects were assigned 1 point for each basic fact mentioned in their problem description. Then the points were totalled to obtain a correspondence score.

Pearson product-moment correlation coefficients were computed for each of the problem types to assess the interrater reliability of the Correspondence ratings. For the arrangement problems, $r = +.838$; for the transformation problems, $r = +.712$; and for the inducing structure problems, $r = +.752$. The variable representing Correspondence was constructed by computing the average of the two sets of ratings.

Measure of Coherence

Coherence was coded by counting the number of required solution steps mentioned by the subjects when retrospectively describing their solutions. Subjects were assigned 1 point for each step mentioned. They were also assigned 1 point for each time a solution step was mentioned immediately before or after the "next" solution step. Thus a subject who mentioned two steps in the correct sequence would receive a total of three points for those two solution steps.

Pearson product-moment correlation coefficients between raters were computed for arrangement problems, $r = +.858$; for the transformation problems, $r = +.810$; and for the inducing structure problems, $r = +.883$. The variable representing Coherence was constructed by averaging the two sets of ratings.

Measure of Connection to Other Knowledge

Connection to other knowledge was coded by counting the number of different problems or situations that subjects said were similar to the current problem. Subjects received 1 point for each similarity mentioned.

Pearson product-moment correlation coefficients were computed for each of the problem types to assess interrater reliability of the Connection to other knowledge ratings. For the arrangement problems, $r = +.866$; for the transformation problems, $r = +.811$; and for the inducing structure problems, $r = +.860$. The variable representing Connection to other knowledge was also constructed by averaging the independent ratings together.

Measure of Structural Understanding

The written protocols provided by subjects as they worked on the problems (their work sheets) were binary coded as reflecting the presence or absence of structural understanding. Good structural understanding was defined as evidence that the subjects grasped the principle underlying the solution to a problem: the structure of the relation between the problem and its solution had to be evident from the worksheet.

Interrater reliability was assessed by computing the percent agreement between the raters. For the arrangement problems, the raters had 84.3% agreement (43/51); for the transformation problems, the raters had 92.2% agreement (47/51); and for the inducing structure problems, the raters had 84% agreement (42/50). The variable representing structural understanding was constructed by averaging the ratings assigned by each of the coders.