INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0500
Insight problem solving: An examination of the process and training of insight

Bohannon, Nancy Kimberly, M.A.

Rice University, 1994
RICE UNIVERSITY

INSIGHT PROBLEM SOLVING:
AN EXAMINATION OF THE PROCESS AND TRAINING OF INSIGHT

by
NANCY KIMBERLY BOHANNON

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
MASTER OF ARTS

APPROVED, THESIS COMMITTEE

John Brelsford, Professor, Chair
Psychology

David Lane, Associate Professor
Psychology

James Pomerantz, Professor
Psychology

Houston, Texas
May, 1994
ABSTRACT

Insight Problem Solving:
An Examination of the Process and Training of Insight

by
Nancy Kimberly Bohannon

In two studies, the process of insight as used in problem solving and the training of insight were examined. In Experiment 1, subjects either practiced solving insight problems or studied insight problems before attempting to solve an insight problem where they were able to ask limited questions of the experimenter. In Experiment 2, subjects were either trained on identifying common assumptions or given practice solving insight problems. Solution rates and verbal protocols for four test problems were collected. Neither study provided evidence of the re-organization of subjects' problem knowledge (as modeled by the Pathfinder algorithm) during problem solution. Though type of training did influence the rate and types of questions asked when solving the test problems, it did not facilitate the actual solution of those problems. The definition of insight was discussed as the identification of relevant and irrelevant problem features rather than as a change in problem representation.
Acknowledgments

There are many people whose help was instrumental to the completion of this project. I would first like to thank my two primary advisors, Dr. Nancy Cooke and Dr. John Brelsford, and the rest of my thesis committee, Dr. David Lane and Dr. Jim Pomerantz. Their guidance was often sorely needed and always vastly appreciated. Even more invaluable was the solid support and complete faith of my family: Owen and Linda Bohannon, Kerry, Kelly, Michael, and Rosa Mae Ferrell. Ours is a close and loving family, and I am eternally grateful that I belong to it. Finally, I would like to specifically thank my mother for her indispensable and priceless labor in transcribing over 100 hours of verbal protocols. Without her, I would still be fast-forwarding and rewinding endless miles of tape.

By space, the universe encompasses and swallows me as a particle; by thought, I encompass it.

> Pascal <

Education is not the filling of a pail, but the lighting of a fire.

> William Butler Yeats <
# Table of Contents

1. List of Tables .................................................................................. v
2. List of Figures .............................................................................. vii
3. Introduction ................................................................................... 1
4. Experiment 1 ................................................................................ 19
5. Method.......................................................................................... 20
6. Results.......................................................................................... 26
7. Discussion ..................................................................................... 61
8. Experiment 2 ................................................................................ 64
9. Method.......................................................................................... 65
10. Results......................................................................................... 70
11. Discussion .................................................................................... 94
12. General Discussion ...................................................................... 95
13. References .................................................................................. 104
14. Appendix A ................................................................................ 107
15. Appendix B ................................................................................ 109
16. Appendix C ................................................................................ 112
17. Appendix D ................................................................................ 113
18. Appendix E ................................................................................ 116
19. Appendix F ................................................................................ 121
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table Caption</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1 - The Number and Percentages of Solvers and Nonsolvers in Each Training Condition in Experiment 1</td>
<td>31</td>
</tr>
<tr>
<td>Table 2 - The Number and Percentages of S1 Solvers, S2 Solvers, and Nonsolvers in Each Training Condition in Experiment 1</td>
<td>31</td>
</tr>
<tr>
<td>Table 3 - The Number of Subjects in Each Condition Who Solved the Problem During the Solving Session or in the Ratings Sessions in Experiment 1</td>
<td>33</td>
</tr>
<tr>
<td>Table 4 - Question Types and Examples</td>
<td>41</td>
</tr>
<tr>
<td>Table 5 - Time 1 Ratings' Correlations Within and Between Training Conditions in Experiment 1</td>
<td>55</td>
</tr>
<tr>
<td>Table 6 - Time 1 Ratings' Correlations Within and Between Training Conditions in Experiment 1</td>
<td>56</td>
</tr>
<tr>
<td>Table 7 - Correlations Between Time 1 and Time 2 Ratings for S1 Solvers and Nonsolvers in the Three Training Conditions in Experiment 1</td>
<td>57</td>
</tr>
<tr>
<td>Table 8 - Similarity Scores Between Time 1 and Time 2 Networks for S1 Solvers and Nonsolvers in the Three Training Conditions in Experiment 1</td>
<td>58</td>
</tr>
<tr>
<td>Table 9 - The Distribution of Solvers and Nonsolvers for Each Assumption Test Problem in Experiment 2</td>
<td>73</td>
</tr>
</tbody>
</table>
# List of Tables (Con't.)

**Table Caption**  
**Table 10** - The Distribution of Solvers and Nonsolvers for the Bar Problem in Experiment 2  
Page: 74

**Table 11** - The Mean Proportions of Strategy, Hypothesis-testing, and Information-gathering Questions Asked in Each Condition in Experiment 2  
Page: 79

**Table 12** - The Presence of the Remedy-Surprise Link for Solvers and Nonsolvers in Time 1 and Time 2 Networks in Experiment 2  
Page: 85

**Table 13** - Time 1 Ratings' Correlations Within and Between Training Conditions in Experiment 2  
Page: 87

**Table 14** - Time 2 Ratings' Correlations Within and Between Training Conditions in Experiment 2  
Page: 88

**Table 15** - Similarity Scores Between Time 1 and Time 2 Networks for Solvers and Nonsolvers in the Three Training Conditions in Experiment 2  
Page: 89

**Table 16** - Correlations Between Time 1 and Time 2 Ratings for Solvers and Nonsolvers in the Three Training Conditions in Experiment 2  
Page: 90

**Table 17** - Goal Statement Categories in Experiment 2  
Page: 91

**Table 18** - The Frequency of Occurrence of Five Goal Categories for the First Four Goal Statements in Experiment 2  
Page: 93
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure Caption</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 1 - The nine-dot problem and solution</td>
<td>4</td>
</tr>
<tr>
<td>Fig. 2 - The mutilated checkerboard problem</td>
<td>7</td>
</tr>
<tr>
<td>Fig. 3 - &quot;AHA!&quot; ratings for the ten training problems for Active training solvers, Active training nonsolvers, and Passive training nonsolvers in Experiment 1</td>
<td>28</td>
</tr>
<tr>
<td>Fig. 4 - The total rate of questions asked by solvers and nonsolvers in each training condition in Experiment 1</td>
<td>35</td>
</tr>
<tr>
<td>Fig. 5 - The total rate of questions asked by S1 solvers, S2 solvers, and nonsolvers in each training condition in Experiment 1</td>
<td>37</td>
</tr>
<tr>
<td>Fig. 6 - The Session 1 question rates for S1 solvers, S2 solvers, and nonsolvers in Experiment 1</td>
<td>38</td>
</tr>
<tr>
<td>Fig. 7 - The Session 2 question rates for S2 solvers and nonsolvers in Experiment 1</td>
<td>39</td>
</tr>
<tr>
<td>Fig. 8 - The proportion of Information-gathering questions asked by solvers and nonsolvers across training conditions in Experiment 1</td>
<td>44</td>
</tr>
<tr>
<td>Fig. 9 - The proportions of Strategy, Information-gathering, and Hypothesis-testing questions asked by solvers and nonsolvers in the Active training condition in Experiment 1</td>
<td>45</td>
</tr>
<tr>
<td>Figure Caption</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Fig. 10 - The relative proportions of the four categories of Information-gathering questions across training conditions in Experiment 1</td>
<td>47</td>
</tr>
<tr>
<td>Fig. 11 - The proportion of Water Information-gathering questions asked by solvers and nonsolvers across training conditions in Experiment 1</td>
<td>49</td>
</tr>
<tr>
<td>Fig. 12 - Example Pathfinder networks in Experiment 1</td>
<td>51</td>
</tr>
<tr>
<td>Fig. 13 - The relative proportions of Strategy, Information-gathering, and Hypothesis-testing questions asked in each condition in Experiment 2</td>
<td>80</td>
</tr>
</tbody>
</table>
Insight Problem Solving:
An Examination of the Process and Training of Insight

Problem solving is one of the more ubiquitous human activities that is engaged in practically every day. For many of the problems that humans face, standard strategies and rules for solving have been developed. In general, problem solving proceeds in a step-wise fashion from the perception of the problem to the generation of the solution. Newell and Simon (1972) break the problem solving process down into components. The problem itself presents the task environment, the solver’s perception of that task environment is called the "problem space", and the solver uses strategies to move within the problem space to the solution. Cognitive psychologists have identified several strategies or processes which people use to solve normal problems: for example, trial and error, using sub-goals and working backward, and means-end analysis (Baron, 1988).

However, there are some problems for which these standard formulae do not work. For these problems, there appears to be no standard formula for solving; in fact, solution often appears to occur randomly. When this kind of problem is solved, the solution appears suddenly and is accompanied by a subjective sense of surprise, an "AHA!" feeling (Gardner, 1978). This type of problem has been called an insight problem, and insight has been minimally defined as this AHA! event. In insight problem solving, there appears to be little or no relationship between the search through the problem space and finding the solution. Often when the insight occurs, it involves a sudden comprehension of the
problem's solution and appears to be independent of whatever strategy the solver was previously using to solve the problem. Insight problem solving and the phenomenon of insight itself are two of the more intriguing and poorly understood aspects of human cognition.

According to Gestalt psychology, the elements of both a problem and its solution are present in the solver's perceptual environment (Ellen, 1982). Problem solving is then the process of correctly perceiving this solution. In the case of insight problems, the solution is not immediately perceived. Thus, what is needed to perceive the solution is a change in the perception of the problem. The Gestalt view defines insight problem solving as this moment when "familiar objects [acquire] a place in a new perceptual organization" where they then take on new meanings (Ellen, 1982, p. 324). This reorganization of familiar objects into new meanings occurs as a sudden perceptual change akin to the change in the perception of figure-ground relationships. The AHA! phenomenon follows this change when the relevance of these new meanings of familiar objects is recognized. According to Ellen (1982), what we call insight, equivalent to the AHA! event, simply accompanies the problem solving process; in other words, it is an epi-phenomenon, but it is not a process itself.

The information-processing paradigm posits an explanation of insight problem solving similar to that of the Gestalt view. Rather than claim changes in the perception of the problem, information-processing theories discuss changes in the cognitive representation of the problem. The process of changing a representation may be very different from that of changing a perception. For example, perceptual change seems to be
more of a chance, random event, whereas representations may be affected by new knowledge and problem solving strategies. Montgomery (1988) presents a three-component model of insight problem solving: 1) the restructuring of a mental model of the problem; 2) the examination of the new mental model; 3) the solution or insight arising from the new model. The primary difference between such an information-processing view of insight and the Gestalt view is that the moment of insight is defined as the access of the problem solution and hence, an integral part of the problem solving process. There is no evidence to date that clearly defines the relative roles of problem restructuring, accessing the problem solution, and the subjective experience of insight itself.

Kaplan and Simon (1990) also discuss insight problem solving as a process involving a change in the representation of the problem. They hypothesize that the solver's initial representation of the problem space is often incorrect and does not permit easy access of the problem's solution. This initial representation must change if the solver is to achieve the problem solution. In other words, insight occurs when the problem space or the solver's representation of the task environment changes. One example of such a problem where a change in problem space is needed for solution is the nine-dot problem (Figure 1). In the nine-dot problem, subjects are presented with nine dots arranged in a 3 x 3 square and are instructed to connect all the dots using only four straight lines, without lifting the pencil from the paper.
The usual initial attempts at solving this problem resemble attempts at solving dot-to-dot problems (Weisberg & Alba, 1981). Thus, the initial problem space is that of a dot-to-dot problem. However, such a representation of the problem will not lead to solution. In dot-to-dot problems, one's lines stay within the perceptual boundary of the dots, but in the nine-dot problem at least one line must leave this perceptual boundary. Until the subject changes his or her problem space to permit lines that leave the boundary of the square, the solution will not be obtained. In fact, very few people solve this problem (Weisberg & Alba, 1981). When solution attempts are examined, most, if not all, of the attempts remain within the boundary of the square. Thus, failure to solve the problem is at least in part due to an incorrect problem space.
Factors Affecting Insight

Research in the area of insight has identified several factors that are related to insight problem solving. Fixation, or the inability to perceive one or more aspects of a problem, is one factor which inhibits insight by preventing the solver from changing his or her representation of the problem. When incubation occurs, however, fixation is overcome. The characteristics of the problem itself also influence problem representation. Providing the solver with hints can influence the solver's problem representation. Finally, training in some problem solving skills may be effective in helping the solver identify the correct problem representation.

Several studies have addressed the issue of incubation (see Kaplan & Davidson, 1988, for a review). The term "incubation" refers to the phenomenon where an insightful solution to a problem is achieved after the problem solving process is interrupted by some other, non-problem solving activities. Smith and Blankenship (1989) hypothesize that incubation effects can occur only when fixation has occurred. They suggest that fixation occurs when "inappropriate information and strategies from memory" (p. 311) block a person's ability to focus on appropriate information and strategies. If an incubation period is then allowed, the inappropriate material is likely to be forgotten, thus allowing access to the appropriate information.

Smith and Blankenship (1989) found that allowing an incubation period of either 5 or 10 minutes increased a subject's likelihood of solving verbal rebuses. In addition, subjects who were allowed an incubation period were less likely to recall a misleading clue that was given when the
test rebus was first presented. The authors concluded that if subjects forgot
the original misleading clue which increased fixation, they were more
likely later to be able to solve the problem they were fixated on. In Smith
and Blankenship (1991), similar incubation effects were found for subjects
completing the Remote Associations Test. Though verbal rebuses may not
be considered insight problems, the Remote Associations Test does involve
insight-related abilities: subjects are asked to think of as many unusual uses
for common objects as they can.

The concepts of fixation and incubation address the issue of changing
problem representations. Insight problems may be so difficult to solve
because inappropriate knowledge and strategies make it difficult to realize
that the original representation of the problem is incorrect. Incubation
provides time for these inappropriate materials to be forgotten. However,
these concepts do not explain how the correct problem representation is
finally identified. In addition, insight does occur when no incubation
period is allowed. Thus, incubation can only partly explain the process of
insight. How inappropriate representations are discarded and correct ones
found when no incubation period is permitted is still unclear.

Kaplan and Simon (1990) investigated the effects of hints on insight
problem solving but used verbal reports from subjects to examine changes
in the problem space more thoroughly. In their study, subjects attempted
to solve the mutilated checkerboard problem. If subjects did not make
adequate progress in solving the problem, the experimenter provided hints
which included disclosing more appropriate problem spaces than the
subject may have been using. In addition, they manipulated the perceptual
characteristics of the problem itself to facilitate access of the correct problem space.

The mutilated checkerboard problem presents a checkerboard with two corner squares (on either diagonal) removed (Figure 2). The subject is asked if he can cover the entire board using pieces like dominoes which cover two squares each. If he cannot cover the board, he must prove that it is impossible to do so. The solution to the problem involves the concept of parity. If the original board has 64 squares (8 x 8) and two diagonally opposite corners are removed, then both of those corners were of the same color. That means that the remaining color now has 2 more squares than the first. The covering dominoes can cover two squares at a time, but no two neighboring squares are of the same color. The leftover two squares, of the color opposite of the removed squares, will always lie in a diagonal; therefore, they cannot be covered by a single domino. Thus, the solution is that it is impossible to cover the board because of the parity of the two square types.

Figure 2

The mutilated checkerboard problem
Kaplan and Simon identified several problem representations that they expected subjects to use (i.e. different covering attempts, mathematical rules, physical manipulation of the board). Via the verbal protocols, they were able to examine the problem representations actually used by the subjects and how these representations were changed. By providing hints, they could influence the problem spaces used. Changing perceptual aspects of the checkerboard also affected the subjects' use of problem spaces. When the appearance of the board emphasized the parity of the two square types (i.e. squares labeled "BREAD" and "BUTTER" versus colored black and white), the subjects' choices of problem spaces were affected and solution was achieved more quickly and more often.

The performance of Kaplan and Simon's subjects followed closely what one would expect if subjects were exploring different problem representations in solving the problem. According to Kaplan and Simon, the most difficult part of the problem for their subjects was identifying the appropriate problem space. Unfortunately, Kaplan and Simon's observations cannot tell us much about how subjects identified the need for a new problem representation or the processes of generating the appropriate new problem representations. Since they provided hints to their subjects, the subjects' choices of representations were not autonomous. However, they were able to show that the characteristics of the problem itself can be a strong influence on which representations subjects initially explore.
Other studies have investigated the effect of providing hints to subjects trying to solve an insight problem. Lockhart, Lamon, and Gick (1988) gave subjects hints to the solutions of some insight problems prior to presenting the problems to the subjects. The hints were presented either in a declarative sentence format or in a puzzle format. For example, consider the following insight problem: "A man who lived in a small town married twenty different women of the same town. All are still living and he never divorced a single one of them. Yet, he broke no law. Can you explain?" The answer is that the man was a clergyman and the twenty women he married, he married to twenty different men. The declarative form of the hint that Lockhart et al. provided was "It made the clergyman happy to marry several people each week." The puzzle form of the same hint was "The man married several people each week because it made him happy." After a few seconds delay, this statement was followed by the word "Clergyman". By separating the two concepts, marrying different women and being a clergyman, this second form of hint resembles the form of an insight more closely than does the declarative format because it initially presents an "inappropriate conception followed by a clue that initiates a process of reconception" (Lockhart, Lamon, & Gick, 1988; p. 37). In other words, the puzzle format was expected to be more effective than the declarative format in facilitating a change in problem representation because it was more similar to the format of the insight problem than was the declarative format.

Though both forms of hints were helpful in later insight problem solving compared to a control group who received no hints, the puzzle
format facilitated later solving more than did the declarative format. Lockhart et al. explain this superior facilitation in terms of the puzzle format's ability to help subjects recognize the relationship between the puzzles and the insight problems. According to the literature on transfer of training, such superiority is to be expected because the surface features of the puzzle hint resembled the problem more than did those of the declarative hint (e.g., Gick & Holyoak, 1980; Holyoak & Koh, 1987). Thus, transfer was facilitated.

Finally, there is some evidence to show that insight can be trained or facilitated. Jacobs and Dominowski (1981) demonstrated that there are practice effects with insight problem solving. In their study, solution times decreased with practice when practice with at least three insight problems was given. However, Jacobs and Dominowski did not provide solving rates with the time data. They concluded that the insight event was not completely problem-specific since it occurred more quickly as a function of practice. Thus, their subjects were learning something during practice that facilitated later insight problem solving. Unfortunately, the data do not describe what they were learning. Lockhart et al. (1988) showed that subjects who received practice on simple puzzles were better at later insight problem solving that were subjects who were given the same information in a non-problem solving format. Again, whatever the subjects had learned from that earlier puzzle experience facilitated insight across a series of insight problems. Thus, something non-problem specific had been learned.

Wicker, Weinstein, Yelich, and Brooks (1978) compared the effects of two forms of skills training on later insight problem solving. The first
training skill emphasized the need to reformulate one's perception of the problem and the need to carefully check assumptions made about the problem. This was accomplished by having subjects work through several insight problems where the solutions were accompanied by explicit descriptions of the structure of the problems and how to go about solving them. The emphasis was placed on not being too narrow when initially formulating the problem, on reformulating the problem, and on checking assumptions. The second type of training instructed subjects to visualize the various problem components. They tested four training conditions: reformulation/assumption checking with visualization, visualization alone, simple practice on insight problems, and simple practice on non-insight problems (control condition). The practice conditions included feedback on the correct solutions to the problems. The dependent measure was the solution rate on 11 insight problems. Results showed that subjects who received both reformulation/assumption checking and visualization training solved more test problems than did subjects in the visualization alone and in the non-insight practice condition (control condition). The subjects in the reformulation/assumption checking and visualization condition did not solve significantly more problems than did subjects in the insight practice with feedback condition. A second study produced similar results: training in reformulation and assumption checking alone (no visualization training) facilitated insight more than did practice on insight problems (without feedback on correct solutions) and control/no training condition.

Unfortunately, there is not sufficient data on how subjects used the training when solving the test problems. The skills of problem
reformulation and assumption checking are very different skills. In these studies, they were confounded. In addition, the insight practice with feedback condition in the first study performed equally well as did the skill training condition. It is not clear just what the subjects learned from the training sessions, how it differed from simple practice, or how that training affected the insight process.

Evidence of Knowledge Re-Organization

Dayton, Durso, and Shepard (1990) defined insight as a sudden change in knowledge organization. This definition is similar to the definition of insight as a change in problem representation (Sternberg & Davidson, 1982; Kaplan & Simon, 1990). They explored this change in depth using the Pathfinder algorithm. Their subjects were trying to solve the following problem:

A man walks into a bar and asks for a glass of water. The bartender pulls a shotgun on the man. The man says "Thank you" and walks out. What missing piece of information would cause the puzzle to make sense?

Relatedness ratings on pairs of key concepts from this problem were collected and submitted to Pathfinder. The Pathfinder algorithm used the subject's relatedness ratings to create a network structure. The network's nodes represent the concepts, the links between nodes represent the connection that the subject perceived between the concepts, and the link weights reflect the strength of those connections.
The Pathfinder algorithm uses proximity data to construct a network of the most efficient paths connecting the nodes. There are two parameters that affect the construction of the network: \( r \) and \( q \). The \( r \) parameter defines the way link weights are summed across a path. For example, \( r = 1 \) indicates that distance between concepts is determined by summing the weights of the links that connect the concepts. In contrast, when \( r = \infty \), the path length is determined by the maximum link weight along the path. The \( q \) parameter limits the number of links allowed in an acceptable path; setting \( q \) to one less than the number of nodes allows paths of any number of links. Varying this parameter affects the density of the final network.

Networks can be constructed from both ordinal and ratio data. Depending on the proximity data supplied, the networks can be undirected (where the weight or value of the link from A to B is the same as from B to A) or directed (the weight or value of a link depends upon which direction you consider). Compared to other modeling techniques such as multidimensional scaling (MDS), Pathfinder networks better represent local relationships among components and they retain some information about the objects that is lost in other techniques. The Pathfinder algorithm has been shown to represent hierarchical relationships and some semantic relationships accurately (see Appendix A for a demonstration of how Pathfinder builds a network; for more detailed descriptions and empirical tests of the Pathfinder algorithm, see Schvaneveldt, 1990).

Dayton et al. (1990), use Pathfinder to construct several group networks representing individual subjects' conceptual organizations of the bartender problem in each of the four following conditions. Subjects in the
Control condition heard the problem and immediately provided the relatedness ratings. They did not attempt to solve the problem, and they did not know the answer. Subjects who tried to solve the problem were allowed to ask yes/no questions for up to two hours. Some subjects solved the problem and then provided the relatedness ratings (Active solvers). Some subjects did not solve the problem in two hours, and they also provided relatedness ratings (Active nonsolvers). Finally, one group of subjects listened to another solver subject's questions. Those subjects, who consequently did not solve the problem themselves, also provided ratings (Passive nonsolvers). Ratings were averaged across subjects in each condition to form group matrices which was then used to construct group networks, rather than individual networks for each subject.

By comparing these four group networks (one for each condition), Dayton et al. could determine that the organization of conceptual knowledge related to the problem was different for solvers and nonsolvers: links between key concepts were present in the solvers' network but were absent in the non-solvers' networks. They concluded that the solution of the problem is accompanied by a change in the solvers' conceptual organization of the problem. The lack of difference in the networks of the active and passive non-solvers indicates that this re-organization was not necessarily a result of active processing of the problem information. Only when the problem was actually solved did the networks differ. Dayton et al.'s data do not show, however, if the solvers' initial approach to the problem or initial use of problem representations differed from that of nonsolvers'. In fact, it is very possible that the initial knowledge
organizations of solvers were different from those of the nonsolvers. If that is the case, then there is no evidence for the re-organization of knowledge at the time of solution.

If reorganization did indeed occur, it is difficult to determine exactly what caused the difference in the conceptual networks between the solvers and non-solvers because there is a confound in this study. The solvers differ from the non-solvers in two ways: they have experienced the process of insight and they know the solution to the problem. Thus, it cannot be determined if the process of insight is responsible for the difference in the networks or if knowledge of the solution itself is responsible. Since Dayton et al. do not discuss the questions that their subjects asked or provide any problem-solving process data, the actual problem spaces used by the subjects cannot be examined. As in Kaplan and Simon (1990), the primary dependent measure in this study was the outcome of insight: whether the problem was solved. Although the Pathfinder networks do provide some information about the solver's and non-solver's conceptual representations of the problem, these networks only illuminate a single point in the insight process. If this type of data could be examined in conjunction with the questions asked by subjects or with think-aloud protocols, the information then collected would provide a much more comprehensive and more descriptive account of the insight process itself.
Limitations of the Insight Literature

In summary, most research in insight problem solving agrees that the insight process involves changes in the representation of the problem or the problem space. This process is also accompanied by a subjective experience, an AHA! event. Several factors have been identified that influence insight, including incubation, training, hints, and the perceptual features of the task environment. However, most of these studies have primarily measured the outcome of the insight process and can provide only limited information about the process itself. Some of the previous research has used techniques such as think-aloud protocols and the Pathfinder algorithm to explore changes in problem representation, but very little data has been collected concerning the process(es) that leads to such a change without the interference of hints or clues. Kaplan and Simon's process data (1990) provides limited information on the process of representation change because they provided their subjects with the correct representation if the subject could not find it on his or her own. Thus, we cannot determine what leads to this change under normal circumstances.

The research on the training of insight is similarly limited: the Jacobs and Dominowski (1981) study, the Lockhart et al (1988) study, and the Wicker et al. (1978) studies did not examine how subjects solved the insight problems, just whether or not subjects solved them and how quickly. Only in the Wicker et al. studies is the training of problem solving skills studied. However, there are additional problems with these studies. In the first Wicker et al. study, the authors did not find any difference in insight facilitation between the reformulation/assumption
checking condition and the condition where subjects practiced solving insight problems and were given feedback on the answers. In the second study, the problem practice condition did not include feedback on answers. In this study, skill training was superior to practice in facilitating insight, but only when the practice did not include feedback. Thus, it is appears that the feedback was an important aspect of the training. There is no conclusive evidence that the skill training provided anything more than simple practice with feedback could.

Wicker et al. did not provide many examples of the training or the test problems, and the few examples they did provide do not appear to be typical insight problems. In addition, the problems used in the Jacob and Dominowski study (1981) were object-use problems only. This type of problem requires the solver to discover a novel use for a common object. The type of insight needed in these problems may not be the same that is required in other insight problems. Lockhart et al. (1988) used verbal insight problems that resemble riddles. Each insight study uses different kinds of insight problems. This diversity reflects a major weakness in the insight literature in general: the lack of a concrete definition of an insight problem. The range of problems that are labeled as insight problems is vast and varied. Some insight problems are spatially based like the nine-dot problem while others may be logic problems, mathematical problems, verbal problems, or even planning problems. A complete theory of insight must be able to account for the insightful solution of all these problem types. The hypothesis that a problem representation change is needed for the solution of the problems can be applied to all these problem types.
However, the type of change needed is different for each problem type. Thus, it is difficult to abstract a common skill or strategy that will apply to more than one type of insight problem. For this reason, evidence that general skill training can facilitate insight is important, both in addressing the many types of insight problems, and in developing a better understanding of insight itself.

Present Research

The present study focuses on three issues. First, does the insight process involve a qualitative change in problem representation? Second, what processes, strategies, or techniques are related to solving insight problems? Finally, can insight be trained? The primary test problem was the Bar problem used by Dayton et al. (1990). To address the first issue, subjects provided relatedness ratings of pairs of the concepts in the bar problem as in Dayton et al. (1990). Pathfinder networks were used to examine the structure of the subjects' problem spaces at two separate points during the problem solving process: at the beginning of the solving session and halfway through the session (Experiment 1) or at the beginning of the solving session and at the end (Experiment 2). The initial networks for solvers and non-solvers are to be compared to determine if the solvers begin the solving process with a different problem representation than do the non-solvers.

Secondly, subjects provided think-aloud protocols as they asked questions while trying to solve the Bar problem. These protocols and the subjects' questions provide a more dynamic view of how people try to
work insight problems. No overt hints were provided in either study. The questions that subjects ask may reveal various approaches or strategies to the problem and the different problem representations that may be used. Together, the two techniques, Pathfinder networks and verbal protocols, should reveal both the changes in subjects' problem representations and the process of that change.

The final focus of the current research was to examine further the possibility of training insight. In the first experiment, two forms of training were tested: practice with feedback similar to the reformulation training in the Wicker et al. (1989) first study and simple instruction without problem solving practice. These conditions also replicated Durso et al.'s (1990) active and passive conditions. In the second experiment, training of a specific skill (assumption checking) was compared to active problem solving practice.

Experiment 1

The purpose of the training conditions in Experiment 1 is to expose the subjects to the structure and nature of insight problems and to provide experience with the strategies needed to solve those problems (and thus facilitate representational change). The training conditions in this experiment contrasted insight problem solving practice with study of insight problems. In the Active training condition, subjects tried to solve a set of verbal insight problems (see Appendix B). If they could not solve a problem within a 2-minute period, the solution was provided along with a
brief description of how the problem misled the subject and how the solution could have been discovered. In the Passive training condition, subjects studied the same problems, solutions, and descriptions but were shown the solution first so that they would not engage in any problem-solving activity. All training subjects were instructed to examine the relationship between the actual solution and the initial problem representation given by the problem. If training is successful, then the subjects should become more proficient at solving the test insight problem compared to control subjects who receive no training.

If Active training subjects solve any of their training problems, they should also be experiencing insight (by current definition, insight is required to solve an insight problem). Passive training subjects do not have this opportunity since they are not allowed to solve the training problems. Passive training subjects are shown the solution to each problem first, and then shown the problem. In this way, the subjects can study the relationships between problem structures and solutions without engaging in the process of problem solving. This condition allows a comparison between training that involves instruction in the need for representational change in solving insight problem (Passive training) and training that involves the actual experience of insight itself (Active training).

**Method**

**Subjects**

The subjects in this study were 45 undergraduate psychology students at Rice University. They received course credit for voluntary participation
in this research. There were approximately equal numbers of males and females in each condition. Subjects were randomly assigned to one of the three conditions with the constraint that there were 15 subjects in each condition.

**Design**

The design was a 3 (Training: None vs. Passive vs. Active) x 2 (Status: Solvers vs. Nonsolvers) between-subjects factorial. Several dependent measures were used: Pathfinder networks, verbal protocols, yes/no questions, subjective insight ratings, and solution rate.

**Materials**

**Control task:** The control subjects rated 40 sentences on a scale measuring sentence complexity. These sentences were unrelated to either the training problems in the other conditions or the test problem.

**Training problems:** A set of verbal training problems was collected from Gardner's 1978 book on insight problems and other studies which have used insight problems. Non-verbal problems such as the nine-dot problem and the mutilated checkerboard problem were not permitted. Many of the problems in this set seemed to be based on single, mistaken assumptions. For example, the following problem is based on the single assumption that the police officer in question is a male: "A police officer had a brother, but the brother had no brother. How is that possible?" If the single assumption is abandoned, then the answer is immediately available. Some problems in the training set did not seem to follow this
structure. For example, the following problem is not based on a single, mistaken assumption: "If you have black socks and brown socks in your drawer mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure you have a pair the same color?" Because the two kinds of problems may involve different solving techniques, the set of training problems was composed of both types. The problems were sorted by 13 graduate students into two categories: assumption-based problems and non-assumption based problems. A problem was categorized as assumption-based if 50% or more of the sorters agreed that it belonged in that category. Problems sorted by less than 50% of the sorters into the non-assumption based category were labeled as non-assumption based problems.

The final set of ten training problems included five of each type. The ten problems and their sources are listed in Appendix B. Each problem was typed on a 5 x 7 inch index card. The solution was typed on the back side of the card. Along with the solution, a brief explanation of the structure and difficulty of the problem was written. For example, on the assumption problems, the statement identified the mistaken assumption that made the problem difficult to solve.

**Test problem:** The bartender problem from Dayton et al. (1990) was used as the test problem. The test problem was also written on an index card, but the solution was not written on the back side of the card.

**Procedure**

Subjects were tested individually. After signing the informed consent form, they were told that the experimenter was interested in
training the ability to solve insight problems. The subject was randomly assigned to one of three conditions and the training phase was begun.

**Control:** During the training phase, Control subjects rated sentences that were unrelated to the insight test problem.

**Passive training:** These subjects were told that the difficulty in solving insight problems is due to the fact that the problem often misleads the solver about what kind of solution is correct. In other words, the wording of the problem usually does not indicate what aspects of the problem are important and how one should think about the solution. These subjects were then shown a series of insight problems; however, the solution to each problem was presented before the problem was presented. This prevented the subjects from engaging in any problem solving activity. They were then instructed to study how different the solution is from what the problem initially appears to ask for. Finally, they rated each problem on a five-point scale for how much of the AHA! type of reaction they think they would have experienced if they had actually solved the problem themselves.

**Active training:** These subjects were also told that the difficulty with insight problems is that the problem often misleads the solver about what kind of solution is correct. However, they were given the training problems and asked to solve each one. Subjects were given two minutes for each problem. If they did not reach the correct solution by the end of the two minutes, they were instructed to turn the card over to read the answer and the explanation of the problem. If the subject solved the problem, he or she was also instructed to read the back of the card so that
every Active training subject read the same explanations that the Passive training subjects did. In all cases, the subject was instructed to observe carefully the discrepancy between their initial problem representation and the actual solution. After completing the ten practice problems, Active training subjects rated each problem on a five point scale for how much of the AHA! experience they had when they either solved the problem or read the problem's solution on the back of the cards.

Test session: After completing the practice problems (Active and Passive training conditions) or the sentence ratings (Control condition), each subject was shown the test problem and asked if he or she had seen the problem before. If the subject recognized the problem as familiar, that subject's data was not included. If the problem was not familiar, then the subject began the first concept ratings session, before beginning to solve the problem. The 14 concepts used in the ratings task are listed in Appendix C. All possible pairs (91) of these 14 concepts were presented on a Macintosh LC computer, and subjects rated how related the two concepts were to each other using a six-point scale. The scale ranged from 1=slightly related to 5=highly related. The sixth option was "Unrelated" for concepts that were perceived as completely unrelated to each other. The ratings were entered on the computer. The order of presentations of the pairs was randomized for each subject and for each time the subject provided the ratings. The order of presentation of items within a pair (A-B vs. B-A) was not counterbalanced since that would have doubled the number of ratings to be made. In other words, the assumption is made that the relationship of A to B is the same as the relationship of B to A. When
this assumption is made, the Pathfinder algorithm does not recognize or produce directional relationships in the networks.

After this first ratings task was completed, the subject was instructed in how to "think aloud" as he or she tried to solve the test problem. A practice think-aloud problem was given (How many windows are in your parents' house?). Each subject was also told that he or she could ask the experimenter yes/no questions about the test problem and the experimenter would provide accurate answers. Subjects were audio-taped using a standard tape recorder.

The solving session was divided into two seven-minute sessions. After the first seven-minute session, the audio-tape was stopped, and the subject was asked to provide the Pathfinder ratings a second time. After the second ratings were collected, a second seven-minute problem solving session was given. At the end of the second problem solving session, subjects who had not solved the problem were given the solution. If a subject solved the problem any time in the first problem solving session, he or she immediately began the second Pathfinder ratings task. Similarly, if the subject solved the problem during the second problem solving session, the session was immediately ended.

After the problem solving sessions were completed, and the answer to the problem given to the nonsolvers, all subjects rated, on a 5-point scale, how much of an AHA! feeling learning or discovering the solution to the bar problem produced.
Results

Subjects could solve the bar problem during either the first or the second problem solving sessions, during the first or the second ratings tasks, or not at all. Subjects were coded as S1 solvers if they solved the problem in the first problem session or in the first ratings session. They were coded as S2 solvers if they solved the problem in the second solving session or the second ratings session. Subjects who did not solve the problem at all were coded as nonsolvers. Thus, subjects who solved the problem in the first solving session (S1 solvers) provided Pathfinder ratings before and after solution. The rest of the subjects provided Pathfinder ratings twice before solution. Therefore, for the Pathfinder analyses, subjects were coded either as S1 solvers or as nonsolvers.

Analyses of the think-aloud protocols and questions examined the data in two ways: 1) S1 solvers vs. S2 solvers vs. nonsolvers and 2) solvers (S1+S2) vs. nonsolvers.

Analysis of Training Problems

As a check on the effectiveness of the Active training condition, solution rates for the ten practice problems were computed. The practice problems were presented in a random order for each subject. The number of problems solved in the first three ordinal positions was compared to the number of problems solved in the last three ordinal positions. Solution rates were higher for the last three problems (72.6%) than for the first three problems (50.3%); this difference was significant, indicating that
performance improved with practice, $F(1,46) = 7.61, p < .009$. The difference in solution rates for assumption-based problems compared with non-assumption based problems was not reliable, $F(1,8) = 0.39, p > .05$.

**Insight Ratings**

Active training subjects rated the practice problems for the amount of AHA! they experienced when they either solved each problem or otherwise learned the solutions. Passive training subjects also rated the amount of AHA! they predicted they might have experienced if they had solved the practice problems. These ratings were compared to examine more closely the subjective experience of insight. Insight ratings were collapsed for solvers, nonsolvers, and Passive training subjects across the practice problem set and submitted to a one-way (Status: Active solver vs. Active nonsolver vs. Passive studiers) ANOVA. The main effect was significant, $F(2,33)=14.174, p < .001$. Active solvers and Passive training subjects gave lower insight ratings ($M=2.92$ and $M=3.1$, respectively) than did the Active nonsolvers ($M=3.97$). In other words, subjects who failed to solve the problem after trying rated the amount of AHA! they experienced when learning the solution higher than did the subjects who successfully solved the problem or who learned the solution without trying to solve the problem. This effect is shown in Figure 3.

One possible way to explain this difference would be to assume that more frustration is experienced when the problem is unsolved. If failing to solve a problem increases frustration and learning the solution eases the frustration, then people who try and succeed in solving the problem will
not experience as much frustration as those who try and fail to solve the problem. Similarly, those subjects who simply read the solution and the problem, and did not attempt to solve the problem, would also experience less frustration, and thus, less AHA!. This hypothesis is only one of many that could explain this difference in subjects AHA! reactions; for example, the effect could be a result of differences in cognitive dissonance between solvers and nonsolvers. The former hypothesis argues an emotional component for the experience of AHA! while the later argues for a cognitive component.

Figure 3
AHA! ratings for the ten training problems for Active training solvers, Active training nonsolvers, and Passive training nonsolvers in Experiment 1

If either hypothesis is correct, then the problems that were more difficult to solve should be given higher AHA! ratings than the easier
problems would be. To test this hypothesis, solution rates were computed for each practice problem; the ten problems were then ranked according to number of subjects who solved them. The three problems that were solved by 55% or less of the subjects were coded as "Hard". Four problems were solved by 63% to 74% of the subjects and were labeled "Medium". The last three problems were solved by more than 82% of the subjects and so were labeled as "Easy". Practice problems were also coded as assumption-based problems or non-assumption based problems. The effects of difficulty level and assumption category were compared in a 2 (Assumption-based: Yes vs. No) X 3 (Difficulty level: Easy vs. Medium vs. Hard) between-subjects ANOVA. The main effect of difficulty level was marginally significant, F(2, 335)=2.89, p=.0572. The Hard problems were given higher AHA! ratings (M=3.62) than were the Medium (M=3.10) or the Easy problems (M=2.89).

Finally, the AHA! ratings for the Bar problem itself were compared in a 3 (Training: Active vs. Passive vs. None) x 2 (Status: Solver [S1+S2] vs. Nonsolver) ANOVA, but neither variable had a significant effect on the AHA! rating. The lack of a reliable difference between solvers and nonsolvers in this rating may be due in part to a ceiling effect. Both solvers and nonsolvers gave the Bar problem very high insight ratings (M=4.0 and M=4.06, respectively).

Of course, there is another possible explanation for these effects on the subjective experience of AHA!: some insights could simply be "bigger" than others. For example, insight problems with trivial solutions might not involve as much insight in their solutions as would less trivial problems.
However, it is not easy to determine exactly what makes a problem trivial or not. Problem difficulty would likely be one component. Unfortunately, we are still left with a circular argument: Trivial problems are solved more easily and with less insight than are non-trivial problems, and problems are trivial if they are solved more easily and with less insight. We need to better define what process or event the AHA! experience reflects in order to better understand its relationship to insight.

Solution Rates Analysis

The primary hypothesis in this study was that active solving of practice insight problems would facilitate later solving of an insight problem. The Chi-Square statistic of the number of subjects who solved the test problem (S1+S2 solvers) in each training condition was not significant, $\chi^2 (2, N=45) = 1.26$, $p > .05$. However, the proportions of solvers in the three training conditions were in the predicted direction. More subjects in the two training conditions solved the problem (50%) than did subjects in the control condition (25%). Also, the Active training condition produced the most solvers (53%). Table 1 shows the number of solvers and nonsolvers in the three training conditions. Table 2 shows the distribution of S1 solvers, S2 solvers, and nonsolvers. This Chi-Square analysis was also non-significant, $\chi^2 (4, N=45) = 1.46$, $p > .05$. 
Table 1
The Number and Percentages of Solvers and Nonsolvers in Each Training Condition in Experiment 1

<table>
<thead>
<tr>
<th>Training Condition</th>
<th>Solvers</th>
<th>Nonsolvers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>8 (53%)</td>
<td>7 (47%)</td>
</tr>
<tr>
<td>Passive</td>
<td>7 (47%)</td>
<td>8 (53%)</td>
</tr>
<tr>
<td>Control</td>
<td>5 (33%)</td>
<td>10 (67%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>20 (45%)</td>
<td>25 (55%)</td>
</tr>
</tbody>
</table>

Table 2
The Number and Percentages of S1 Solvers, S2 Solvers, and Nonsolvers in Each Training Condition in Experiment 1

<table>
<thead>
<tr>
<th>Training Condition</th>
<th>S1 solvers</th>
<th>S2 solvers</th>
<th>Nonsolvers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>3 (20%)</td>
<td>5 (33%)</td>
<td>7 (47%)</td>
</tr>
<tr>
<td>Passive</td>
<td>2 (13%)</td>
<td>5 (33%)</td>
<td>8 (53%)</td>
</tr>
<tr>
<td>Control</td>
<td>2 (13%)</td>
<td>3 (20%)</td>
<td>10 (67%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>7 (15.5%)</td>
<td>13 (28.9%)</td>
<td>25 (55.6%)</td>
</tr>
</tbody>
</table>

Recall that subjects engaged in two separate concept rating sessions. Several subjects actually solved the problem during one of these sessions rather than during a problem solving session. One Active training subject solved the problem in the first rating session, and two Control subjects solved the problem during the second ratings session. However, all seven
Passive training subjects who solved the problem did so during a ratings session: two subjects in the first and five in the second session. The ratings session were a potential source of very helpful hints to the problems. Some of the concept pairs that subjects saw and rated included Glass of Water - Remedy, Shotgun - Surprise, Glass of Water - Relief, and Paper Bag - Remedy. For several subjects, these pairs provided help in solving the problem. In addition, the second ratings session (where more subjects solved the problem than in the first ratings session) provided a period of non-problem solving activity that could serve as an incubation period. The ability to utilize the hints available in the ratings sessions is discussed below.

These ten subjects who solved the problem during a ratings session were obviously making connections during those sessions that other subjects were not. They may have been engaging in a different form of problem solving activity than were other subjects. It seems curious that all of the Passive training condition solvers discovered the solution to the bartender problem during one of the rating sessions, whereas only one or two subjects in the other conditions did so. A Chi-square analysis of the number of problem solving session solvers and ratings session solvers in the three training conditions did provide a significant result, $\chi^2 (2, N=20) =11.7, p<.003$. The frequency table for this analysis is listed in Table 3.
Table 3
The Number of Subjects in Each Condition Who Solved the Problem During the Solving Sessions or the Ratings Sessions in Experiment 1

<table>
<thead>
<tr>
<th>Training Condition</th>
<th>Ratings Session</th>
<th>Solving Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Passive</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Control</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

To look at the problem solving success of those who did not solve the problem during a ratings session, a Chi-square analysis was performed on only the number of subjects in each training condition who solved the problem during one of the problem solving sessions. This analysis produced a significant result, $\chi^2 (2, N=45) = 9.51, p<.01$. More subjects in the Active training condition solved the problem during a solving session (7 out of 15) than did subjects in the Passive training (0 out of 15) or the Control conditions (3 out of 15).

There appear to be two problem solving types occurring in this study: the problem solving activity which uses information gained during the problem solving sessions, and the activity that utilizes the information in the ratings tasks. The different training conditions did influence these two activities differently: Active training subjects were more likely to solve the problem during a problem solving session and Passive training subjects were more likely to solve the problem during the ratings sessions.
Possible differences in the two activities of active problem solving and providing relatedness ratings are discussed below.

Considering the results of these and other analyses discussed below, it seems likely that the non-significance of the overall number of problem solvers in each condition is in part due to a lack of sufficient power in this study. Each training condition had only 15 subjects and the cell sizes are fairly small in each analysis; therefore, a larger sample size may be needed to demonstrate a significant general training effect.

Questions Analyses

Question Rates: The next analysis examined the rate at which subjects in each condition asked questions of the experimenter. The bar problem is not likely to be solved with just the information provided in the problem statement. Additional information can be acquired by asking the experimenter questions; it is assumed that the more questions the subject asks, the more likely that subject should be to solve the problem. Asking questions of the experimenter also reflects at least some kind of problem solving activity that can easily be measured. The total number of questions asked by each subject was tallied, and a ratio was computed by dividing the total number of questions by the amount of time needed to solve the problem. Nonsolvers were scored using 14 minutes, the maximum amount of time available. A 3 (Training: Active vs. Passive vs. None) x 2 (Status: Solver [S1+S2] vs. Nonsolver) ANOVA was performed on the ratios. Neither main effect was significant, but the interaction was marginally significant, \( F(2,39) = 3.19, p = 0.0521 \). The interaction derives from the
fact that of the subjects who solved the problem, both Active and Passive training subjects asked more questions per minute than did the Control subjects. For nonsolvers, the reverse effect is true: Control subjects asked more questions per minute than did the Active or Passive training subjects. Figure 4 demonstrates this interaction. For trained subjects (both Active and Passive training), asking more questions leads to solving the problem; for subjects with no training, asking more questions does not necessarily lead to problem solution. It appears that the questions being asked by the trained subjects may be different from the questions asked by the control subjects.

Figure 4
The total rate of questions asked by solvers and nonsolvers in each training condition in Experiment 1

This interaction was explored further by separating subjects into the three outcome groups: S1, subjects who solved the problem in the first
session; S2, subjects who solved the problem in the second session; and nonsolvers, subjects who did not solve the problem at all. A 3 (Training: Active vs. Passive vs. None) x 3 (Status: S1 vs. S2 vs. Nonsolver) ANOVA was performed on overall question rates. In this analysis, both main effects were significant: a main effect of outcome status, F(2,36)=17.3, p<0.001, and a main effect of training condition, F(2,36)=4.8, p<0.02. In addition, the interaction between outcome and training was significant, F(4,36)=9.62, p<0.001. Specifically, S1 subjects asked the most questions per minute overall, while S2 and nonsolver subjects asked questions at a slower rate. Thus, when subjects began the first solving session with many questions, they were more likely to solve the problem than were subjects who did not ask as many questions.

Figure 5 demonstrates the interaction between training and outcome status for question rates. The effect of training was present only for those subjects who solved the problem in the first session: both Active and Passive training S1 subjects asked questions at a higher rate than did Control S1 subjects. Otherwise, all conditions had equivalent question rates. Thus, the primary effect of training conditions appears for subjects who solve the problem in the initial problem solving session.
Since the previous analyses indicate that the overall question rates differ between S1 solvers (who only participated in the first solving session) and the remaining subjects, the question rates for each solving session were explored separately. Question rates were computed for each solving session in the same manner as the overall question rates were computed. Two 3 (Training: Active vs. Passive vs. None) X 3 (Status: S1 solver vs. S2 solver vs. Nonsolver) ANOVAs were computed, one for each solving session. For the first solving session, the main effects of solution status \( [F(2,36)=4.48, p<0.02] \) and training condition \( [F(2,36)=20.35, p<0.001] \) and their interaction were significant \( [F(4,36)=10.79, p<.001] \). The interaction is illustrated in Figure 6. Again, we see that Trained S1
solvers asked significantly more questions in that first solving session than did the Control S1 solvers. For the S2 solvers and the nonsolvers, the three conditions were about equal in their question rates.

**Figure 6**
The Session 1 question rates for S1 solvers, S2 solvers, and nonsolvers in Experiment 1

In the second solving session analysis, S1 solvers were omitted (they did not have a second solving session). Again the interaction between status and training condition was significant, F(2,30)=3.75, p<0.04. In this session, S2 solvers asked more questions (M rate=6.82) than did nonsolvers (M rate=3.35), but only for the Passive training condition. For the Active training and Control conditions, S2 solvers and nonsolvers asked questions at approximately equal rates. This interaction is illustrated in Figure 7.
The analyses of question rates in each session and overall indicate that, in general, asking more questions facilitates solving the problem, but only for the trained subjects. Trained subjects appear to begin the solving session with more questions than do the control subjects and their questions are more effective than are those of the control subjects since the additional questions lead to solution only for the trained subjects. Thus, it appears that the kind of questions being asked by subjects in the different conditions are not equally effective. It is important to remember that these data do not necessarily indicate a causal relationship between asking questions and solving the problem. It may be simply that subjects who are better problem solvers also ask many questions, not necessarily that asking many questions makes one a better problem solver.
**Question Types:** To examine the effectiveness of the questions subjects asked, it is first necessary to determine what kind of questions are being asked and which kind should be more effective than others. There are three aspects of the Bar problem that are good targets for questioning. These aspects or components concern the three primary actions in the problem and provide ample domains for searching for the solution: 1) Water - Why did the man want a glass of water? 2) Shotgun - Why did the bartender point a shotgun at the man? and 3) Thank you - Why did the man say thank you when he left? Although subjects could not ask these specific questions, their yes/no questions often reflected interest in these three question domains.

After examining the questions asked by subjects, three types of questions were determined: Hypothesis-testing questions, Information-gathering questions, and Strategy questions. These question categories are described in Table 4.
Table 4

Questions Types and Examples

<table>
<thead>
<tr>
<th>QUESTION TYPE</th>
<th>DESCRIPTION</th>
<th>DOMAIN EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis-testing</td>
<td>Pose direct and explicit potential answers to the question domains.</td>
<td>Is the man thirsty? (Water) Was the bartender really busy and the guy started a fight? (Global)</td>
</tr>
<tr>
<td>Information-gathering</td>
<td>Seek additional information without posing solutions.</td>
<td>Did the man thank the bartender? (Thank you) Was it a normal bar? (Other)</td>
</tr>
<tr>
<td>Strategy</td>
<td>Ask about ways to solve the problem, relevance of aspects of the problem.</td>
<td>Are the words in the ratings task important? Does it matter what kind of job the man has?</td>
</tr>
</tbody>
</table>

Each of the subjects' questions were coded as Hypothesis-testing, Information-gathering, or Strategy questions, and could refer to the Water, the Shotgun, or the Thank-You component of the problem. Occasionally, Hypothesis-testing questions referred to more than the water, the bartender, and the thank-you components of the problem. This type of Hypothesis-testing question posited detailed, scenario-like possible solutions to the entire problem, and were coded as Hypothesis-Global questions. Some Hypothesis-testing and Information-gathering questions did not fit into any domain and so were coded as Other.
For the analysis of type of question, a subset of subject protocols was compiled. A sample of 18 subjects, three solvers (S1+S2) and three nonsolvers from each condition, was randomly selected. Subjects who solved the problem in less than 30 seconds were not eligible for this sample because they did not provide enough question data to examine. There were three such subjects (one in the Active training condition and two in the Passive training condition) who solved the problem so fast that there were too few questions to analyze.

The number of Hypothesis-testing, Information-gathering questions, and Strategy questions were tabulated for each subject in the sample. In addition, the Hypothesis-testing questions and the Information-gathering questions were coded as Water, Shotgun, Thank You, or Other. No Hypothesis-testing/Global questions were found in the subset of 18 transcripts. The numbers of Hypothesis-testing, Information-gathering, and Strategy questions were then divided by the total number of questions asked by each subject and multiplied by 100, resulting in three question-type proportions for each subject. Within the Hypothesis-testing type and the Information-gathering type, proportions of questions dealing with the three problem components (Water, Shotgun, Thank You) were also calculated by dividing the number of component questions by the total number of questions of that type asked. For example, one proportion for one subjects was calculated as follows: (Number of Information-gathering-Water questions /Total number of Information-gathering questions)*100.

A 3 (Training: Active vs. Passive vs. None) X 2 (Status: Solver (S1+S2) vs. Nonsolver) X 3 (Type: Hypothesis-testing vs. Information-
gathering vs. Strategy) mixed-factorial ANOVA was performed. The repeated measure was the proportions of questions belonging to each question type. The main effect of Question type was significant, 
\[ F(2,24)=53.83, \ p<0.001. \] Most of the questions asked were Information-gathering questions (51.93%), followed by Hypothesis-testing questions (32.96%). Strategy questions were least often asked (11.09%). Means comparisons indicate that each of the three proportions is reliably different from the remaining two.

The 3-way interaction of Question Type, Training condition, and Status was also significant, 
\[ F(4,24)=3.57, \ p=.0374. \] This complex interaction is better understood by conducting separate analyses for each question type. Thus, three 3 (Training: Active vs. Passive vs. None) X 2 (Status: Solver (S1+S2) vs. Nonsolver) ANOVAs were performed, one for each question type. The proportion of questions that were Hypothesis-testing was not different across training conditions or solution status. However, there were differences in the proportion of Information-gathering questions and Strategy questions asked across all conditions.

First of all, Active training subjects asked more Strategy questions \( (M=18.2\%) \) than did Passive training subjects \( (M=5.2\%) \) or Control subjects \( (M=9.9\%) \), 
\[ F(2,12)=8.27, \ p<0.01. \] Secondly, the interaction between training condition and Status was significant for the proportions of Information-gathering questions asked, 
\[ F(2,12)=5.79, \ p<0.02. \] This interaction is shown in Figure 8. In the Active training condition, more Information-gathering questions were asked by solvers than by nonsolvers. The reverse is true for the Passive training and the Control conditions.
Apparently, subjects in the Active training condition who asked Information-gathering questions were more likely to solve the problem than were Passive training and Control conditions subjects who asked more Information-gathering questions.

Figure 8
The proportion of Information-gathering questions asked by solvers and nonsolvers across training conditions in Experiment 1

Another way of looking at this complex, three-way interaction is by conditions. Thus, the proportions of each type of questions were submitted to three 2 (Status: Solver (S1+S2) vs. Nonsolver) x 3 (Type: Hypothesis-testing vs. Information-gathering vs. Strategy) mixed-factorial ANOVAs, one for each training condition. The variable Question Type served as a repeated measure. For the Active training condition, the main effect of
question type was significant, $F(2,8)=9.96, p<0.007$. Active training subjects asked fewer Strategy questions ($M=18.20\%$) than Information-gathering ($M=45.21\%$) or Hypothesis-testing questions ($M=32.75\%$). The Information-gathering and Hypothesis-testing proportions were not significantly different from each other. The Status X Type interaction was marginally significant, $F(2,8)=4.38, p=.052$. This interaction is depicted in Figure 9.

Figure 9
The proportions of Strategy, Information-gathering, and Hypothesis-testing questions asked by solvers and nonsolvers in the Active training condition in Experiment 1

The primary source of the interaction depends from the proportions of Strategy questions are similar for solvers ($M=17.85\%$) and nonsolvers
(M=18.55%), but not the proportions for Information-gathering and Hypothesis-testing questions. Solvers asked more Information-gathering questions (M=53.87%) than did nonsolvers (M=36.55%). On the other hand, nonsolvers tended to ask more Hypothesis-testing questions (M=42.01%) than did solvers (M=23.49%).

For the Passive training condition, only the main effect of question type was significant, F(2,8)=28.63, p<0.001. Strategy questions were asked least often (M=5.01%), Hypothesis-testing questions next frequently (M=36.41%), and Information-gathering questions the most often (M=55.29%). The proportions for Hypothesis-testing questions and Information-gathering questions are reliably different, F(1,8)=7.98, p<0.03.

Finally, in the Control condition, again only the main effect of question type was significant, F(2,8)=17.79, p<0.01. The same pattern exists for this condition: Strategy questions asked least often (M=9.88%), with Hypothesis-testing questions next (M=29.71%), and Information-gathering questions most (M=55.28%).

Information-gathering questions were further examined by comparing the proportions of Information-gathering questions concerning the Water, Shotgun, Thank You, and Other categories described above. In a 3 (Training: Active vs. Passive vs. None) X 2 (Status: Solver (S1+S2) vs. Nonsolver) X 4 (Information Type: Water vs. Shotgun vs. Thank You vs. Other) mixed-factorial ANOVA, the main effect of Information Type and the Information Type X Condition interaction were significant; F(3,36)=34.00, p<0.001, and F(6,36)=2.88, p<0.03, respectively. Most
Information-gathering questions belonged to the Other category (54.65%) while the smallest category was the Thank You Information questions (1.98%). The Water and the Shotgun categories of Information-gathering questions were approximately equal; 21.8% and 19.17%, respectively.

The pattern across the categories was similar for the three training conditions, but the absolute values differed. The Information Type X Training interaction is shown in Figure 10.

Figure 10
The relative proportions of the four categories of Information-gathering questions across training conditions in Experiment 1

Subjects in the three training conditions asked about the same proportion of Thank You questions. Control subjects asked significantly
more Other questions than did subjects in the two training conditions, whereas subjects in the two training conditions asked more Water and Shotgun questions. The locus of the interaction lies in the fact that the Control subjects asked fewer Water and Shotgun questions but more Other questions than did subjects in either training condition. There was no difference in proportions across categories for the two training conditions. Therefore, Control subjects spent more time asking Other questions and less time asking Water and Shotgun questions than did the Active and Passive training subjects.

Finally, the four categories of Information-gathering questions were examined in four separate 3 (Training: Active vs. Passive vs. None) X 2 (Status: Solver (S1+S2) vs. Nonsolver) between-subjects ANOVAs, one per category. The analysis for the Water category revealed both significant main effects and a significant interaction. In general, trained subjects asked more Water questions (Active M=25.19%, Passive M=27.66%) than did Control subjects (M=15.55%), F(2,12)=4.12, p<0.05. However, only in the Passive training condition did solvers ask more Water questions than nonsolvers did. In the Active training and Control conditions, solvers and nonsolvers asked an equivalent numbers of Water questions. This interaction is shown in Figure 11. None of the remaining category analyses were significant.
The data indicate that, in general, there is some sort of relationship between the number and kinds of questions asked and being able to solve the bar problem. Though it is not possible to determine the nature and direction of that relationship here, it would be possible to test this relationship in a more controlled study where the number and kind of questions to be asked could be experimentally limited.

Pathfinder Analyses - Evidence for Reconstruction

Relatedness ratings for the 14 concepts in the problem were collected twice. Subjects gave Time 1 ratings at the beginning of the first solving session. Time 2 ratings were given at the end of the first solving session or when the problem was solved, whichever came first. Thus, subjects who
solved the problem in the first session gave relatedness ratings before and after solving the problem (S1 solvers). Subjects who solved the problem in the second session or who did not solve the problem at all provided both sets of their ratings prior to solving the problem (nonsolvers). All ratings were submitted to the Pathfinder algorithm with $r=\infty$ and $q=13$, and two networks were constructed for each subject, one for Time 1 and one for Time 2. Comparison of networks across conditions and across time should reveal any concept re-organization that may occur.

Dayton et al. (1990) constructed group networks for their four conditions. However, the use of averaged ratings to create these group networks assumes that the ratings provided by each subject were determined by similar scales. In other words, the authors assume that a relatedness ratings of "3" from each subject indicated the same degree of relationship. When the proximity data collected (here, relatedness ratings) is from a fairly objective measurement, then averaging ratings to form group networks makes sense. However, this assumption cannot be made lightly for the relatedness ratings provided here. There is no evidence that each subject is using the rating scale in an equivalent manner, especially since problem perception may be a very individualized process. Thus, the averaged group networks used in Dayton et al.'s study may not have been appropriate; individual differences among subjects are washed away in such a compilation. Therefore, the rating data collected in the current study was used to form individual networks for each subject. No group networks were constructed. Thus, differences among the training conditions were addressed by comparing within-subject network
differences for subjects in each condition. For examples of individual Pathfinder networks, Figure 12 shows a Time 1 and a Time 2 Pathfinder network for a single Bar problem solver.

Figure 12

Example Pathfinder networks

Control Condition Subject, Time 1 Network

Control Condition Subject, Time 2 Network
In Dayton et al.’s (1990) analysis of their subjects' networks, they focused on the presence and absence of specific links. Their data showed that the group network for subjects who solved the problem contained the links Glass of Water-Remedy and Surprise-Remedy, but the group networks for nonsolvers did not have these links. Dayton et al. proposed that the restructuring of the subject’s concepts occurred during the process of actively solving the problem and was responsible for these differences in links.

In an attempt to replicate Dayton et al.’s results, individual Pathfinder networks for Time 1 and Time 2 were examined for each subject. The presence or absence of the Glass of Water-Remedy link was counted for each Time 2 network. The results of a Chi-square analysis were not significant, $\chi^2 (1, N=45) =0.49$, $p >.05$. The number of times the link was present in individual Time 2 networks was not reliably different for the two groups, S1 solvers and nonsolvers. The presence of the Surprise-Remedy link was also counted for each Time 2 network and the totals analyzed. This Chi-square also failed to show a significant difference across groups, $\chi^2 (1, N=45) =1.51$, $p >.05$. Again, the presence of this link did not differentiate between solvers and nonsolvers.

Correlations between individual Time 1 and Time 2 relatedness ratings were compared for S1 solvers and nonsolvers. If problem reconstruction is occurring during insight, then the correlations between Time 1 and Time 2 should be less for solvers than for nonsolvers because the relationships among the concepts should change as the problem representation does. The average correlation between Time 1 and Time 2
ratings for S1 solvers tended to be lower ($M=0.87$) than the average correlation for nonsolvers ($M=0.96$); however, this difference was not significant.

A similar pattern was found when comparing individual Time 1 and Time 2 network similarity scores. The average similarity score was lower for S1 solvers ($M=0.64$) than for nonsolvers ($M=0.86$). Again, this difference was not reliable. However, the trends are in the right direction if reconstruction is occurring for solvers.

There is one difficulty in interpreting these Pathfinder data, specifically with the Pathfinder data from the Passive training condition. Recall that some of the subjects who solved the problem did so during the rating periods. At what point during each rating period that solution occurred was not, however, recorded. Thus, we cannot determine if the ratings provided by these solvers were actually made prior to or after solution. Because of this ambiguity, the ratings correlations and network similarity analyses were computed a second time without the data from these subjects. Recall also, that all of the Passive training condition subjects solved the problem during a ratings session, and thus, all were excluded from the new analyses. The results were the same as those above: neither the correlations between ratings at Time 1 and Time 2 nor the network similarities between Time 1 and Time 2 were significantly different for S1 solvers and nonsolvers.

The lack of agreement between these Pathfinder results and those for the Dayton et al. (1990) study may be due to two factors. First of all, there are only seven S1 solvers (three of which solved the problem during a
ratings session), and this number may be too small to reveal significant differences in this study. Secondly, subjects were not explicitly instructed to provide their ratings in the context of the Bar problem instead of from their general world knowledge. The Pathfinder analyses is supposed to measure problem representation, not general knowledge representation. Therefore, if subjects were making relatedness ratings from general knowledge instead of from the context of the problem, then not much change should be expected in during the course of solving the problem.

Pathfinder Analyses - Effects of Training

Pathfinder ratings and networks may reflect effects of the training conditions. If subjects are learning something from the training tasks that might influence their ability to solve the insight problem, then subjects' interpretations of the problem would be more similar when the subjects belong to the same training condition than when they belong to different training conditions. This would be reflected in the intercorrelations among subjects within each training condition. Specifically, correlations among subjects' ratings within each training conditions should be higher than correlations across training conditions.

Time 1 ratings from subjects in each of the three training conditions' were correlated with other subjects' ratings within and across conditions. Mean correlations were computed for six groups (Active-Active, Active-Passive, Active-Control, Passive-Passive, Passive-Control, and Control-Control). Table 5 shows these correlations.
Table 5
Time 1 Ratings' Correlations Within and Between Training Conditions
in Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Active</th>
<th>Passive</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>0.466</td>
<td>0.482</td>
<td>0.452</td>
</tr>
<tr>
<td>Passive</td>
<td></td>
<td>0.480</td>
<td>0.456</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td>0.434</td>
</tr>
</tbody>
</table>

The within-group correlations for the Active Condition (.466) and for the Passive Condition (.48) are slightly higher than that for the Control condition (.434). Thus, there may be more internal consistency among the ratings made by subjects in the two training conditions than there is by subjects in the control condition. The correlation between the Active Condition and the Passive Condition (.482) is also slightly higher that those between the Control Condition and either training condition (Active-Control=.452; Passive-Control=.456). This pattern of inter- and intra-group ratings correlations suggests that the training conditions may have influenced the subjects' initial perceptions of the problem. One possible explanation is that the training groups learned something from the training they received, something that led to a more consistent representation of the problem concepts than the Control subjects showed.
A similar examination was conducted for Time 2 ratings revealing the same pattern seen in the Time 1 ratings. Table 6 shows that the correlations within and among the training groups at Time 2 are higher than those correlations involving subjects in the control group. Trained subjects began the solving process with a different knowledge organization than the control subjects did, and this difference was maintained throughout the solving process.

Table 6
Time 2 Ratings' Correlations Within and Between Training Conditions in Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Active</th>
<th>Passive</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>0.434</td>
<td>0.469</td>
<td>0.426</td>
</tr>
<tr>
<td>Passive</td>
<td></td>
<td>0.472</td>
<td>0.443</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td>0.410</td>
</tr>
</tbody>
</table>

To examine the correlations between Time 1 and Time 2 ratings for individual subjects in the different training conditions, correlations between Time 1 and Time 2 ratings were computed for each subject and submitted to a 3 (Training: Active vs. Passive vs. None) X 2 (Status: S1 solver vs. Nonsolver) ANOVA. The mean correlations are shown in Table 7. According to these numbers, the least correlation between Time 1 and Time 2 ratings occurred for S1 solvers in both training (Active and Passive) groups while nonsolvers and all Control subjects showed high
correlations, indicating less change. However, these differences were not reliable.

Table 7
Correlations Between Time 1 and Time 2 Ratings for S1 Solvers and Nonsolvers in the Three Training Conditions in Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Active</th>
<th>Passive</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 solvers</td>
<td>0.766</td>
<td>0.7922</td>
<td>0.8757</td>
</tr>
<tr>
<td>Nonsolvers (S2 solvers + nonsolvers)</td>
<td>0.9317</td>
<td>0.9393</td>
<td>0.9243</td>
</tr>
</tbody>
</table>

Similar networks at Time 1 and Time 2 would also indicate less change than would dissimilar networks. Individual similarity scores were derived for each subject and submitted to a 3 (Training: Active vs. Passive vs. None) X 2 (Status: S1 solver vs. Nonsolver) ANOVA. Again, there were no significant effects. However, the pattern of similarity scores is similar to that of correlations. The overall similarity scores for the six groups are shown in Table 8. The similarity scores are lower for the two training groups than for the control group.
Table 8
Similarity Scores Between Time 1 and Time 2 Networks for S1 Solvers
and Nonsolvers in the Three Training Conditions in Experiment 1

<table>
<thead>
<tr>
<th>Status</th>
<th>Active</th>
<th>Passive</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 solvers</td>
<td>0.4857</td>
<td>0.2258</td>
<td>0.6568</td>
</tr>
<tr>
<td>Nonsolvers (S2 solvers + nonsolvers)</td>
<td>0.6774</td>
<td>0.7976</td>
<td>0.5679</td>
</tr>
</tbody>
</table>

Protocol Analysis

The protocol analyses were conducted using the subset of 18 subjects used in earlier analyses (3 solvers and 3 nonsolvers from each training condition). There was a great variability in the think-aloud portions of the data. Some subjects were more at ease with thinking aloud than others, and thus were more verbal. Some subjects found it harder than others to ask the experimenter questions, even when reminded of this resource. Other subjects simply did not recognize the fact that they could learn additional information about the problem by asking the experimenter questions even though they were instructed to do so. Although all subjects were encouraged to keep talking during the session and reminded that they could ask questions, some subjects were patently better at these skills than others.

Logically, it appears that the best way to solve the problem is to determine why the man wanted a glass of water. However, this issue was
not commonly explored in the data. In fact, the question analyses showed that the shotgun issue was questioned more often than was the water issue. In addition, the shotgun issue appeared to be the first issue explored in the think-aloud data. To test this apparent attraction of the shotgun issue, the initial statements made in each of the 18 protocols were examined. Usually, the subject began trying to solve the problem by reading the problem or portions of the problem aloud. After the recitation, the subject often began exploring one of the three primary components of the problem described above (Why did the man want a glass of water? Why did the bartender pull out a shotgun? Why did the man say thank you and leave?). The number of subjects who mentioned each component first was totaled. Only four of the 18 subjects mentioned the issue of the water first (22.2%), one mentioned the thank you issue first (5.6%), and 13 referred to the shotgun first (72.3%). This distribution is significantly different for the three problem components, $\chi^2 (2, N=18) = 42.83$, p<0.001. Thus most subjects began their first solving session with the issue of the shotgun. The distribution of the four subjects who began with the issue of the water was equally spread across solvers and nonsolvers and across training conditions.

There are other strategies for solving the problem that subjects rarely used. Specifically, they did not recognize the relevance of the ratings concepts even though they were explicitly told that the concepts are related to the Bar problem. Except for the concepts "Friendly" and "Loaded", subjects rarely referred to the ratings list in the first solving session. Only six of the 18 subjects did so in the first solving session (33.3%). Four of these subjects later solved the problem. During the
second solving session, when they had completed the second ratings task, subjects were more likely to notice the relevance of the ratings concepts. Eight subjects (44.4%) began the second session with immediate questions about the ratings concepts; however, seven of these eight did not solve the problem. Four subjects never referred directly to the ratings concepts at all; surprisingly, these four subjects did eventually solve the problem.

Perhaps asking questions about the ratings concepts did not facilitate solving the problem because, when subjects did mention the ratings concepts directly, they usually referred to the concrete concepts that were not in the problem itself: pretzels, TV, and paper bag. Pretzels and TV are unrelated to the solution. While paper bag is highly related to the solution, the relationship is not obvious or easy to determine. Fourteen of the 18 subjects referred to the ratings concepts; of these, ten mentioned TV (71.4%), nine mentioned pretzels (64.3%), and six mentioned paper bag (42.8%).

For the three abstract concepts that were more related to the solution (remedy, relief, and surprise), references were even more scarce. Only three subjects (out of 14) mentioned remedy (21.4%), six mentioned relief (42.8%), and four mentioned surprise (28.5%). It appears that even when considering the ratings concepts, the subjects were likely to focus on the wrong, or at least less useful, concepts.

Finally, there is the issue of memory load. The subjects in this experiment did not have pen and paper to help keep track of their questions and the answers. Many questions were repeated, answers were often forgotten, and sometimes answers were recalled incorrectly. It seems that
subjects were not very good at keeping track of the information they were learning, especially between the two solving sessions. Perhaps because of the memory difficulty, subjects were not good at assimilating the information they were learning. For example, if a subject learned in the first session that the man was not thirsty, he or she was unlikely to remember this in the second session when he or she learned that the water was a remedy. Thus, it would be very difficult for that subject to deduce that the water was a remedy for something other than thirst.

Experiment 1 Discussion

The analyses of subjects' performance when trying to solve the practice problems (Active training) replicated earlier studies by demonstrating that the likelihood of solving a problem increased with practice. Therefore, there is some component or components of the insight process that can be facilitated by practice.

Although there were no significant differences in the Bar problem solution rates for the three training conditions, further analyses of questioning patterns and knowledge structures indicated some other effects of training on the rate and types of questions asked. It is possible that the non-significant effect of the training variable on solution rates was due to a lack of sufficient power in this experiment. Each training condition had only 15 subjects; a larger sample size may be needed to demonstrate a significant training effect. In addition, the type of training given in this study may not have been completely appropriate for solving the Bar
problem. The practice problems given as training may not have had as much in common with the Bar problem as was expected. Half of the practice problems had been previously categorized as assumption-based problems: that is, the problems were easily solved if a single inaccurate assumption were corrected. In trying to solve the Bar problem subjects do make incorrect assumptions, but the solution requires more than the simple correction of these assumptions. The Bar problem is much more difficult than the training problems and may require different solving strategies than what the subjects were learning during training. Although previous research shows that there are practice effects for some kinds of insight problems (Jacobs & Dominowski, 1981), the kind of problems used in training for this study may not have been sufficiently compatible with the Bar problem. The issues of sample size and training-test problem compatibility were addressed in Experiment 2.

The analyses of questions asked during the solving sessions also provided evidence of training effects. Training influenced the number of questions asked and the kind of questions asked, especially for the first solving session. It also affected how the subjects approached the problem-solving task in general; active training subjects asked more Strategy questions to help direct their solution attempts. The Pathfinder data also showed evidence for training effects in subjects' initial perception of the Bar problem. However, for the reasons discussed above, training did not lead to significant differences in solution rates.

The lack of conclusive data in the Pathfinder analyses for insight as restructuring may likely be due to the subjects' tendency to rate the
concepts from their general world knowledge rather than providing ratings restricted to the problem itself. Again, this problem is addressed in the second experiment.

However, there are other issues concerning the use of the Pathfinder algorithm that need to be discussed. First, providing Pathfinder ratings was a time-consuming process for subjects. Subjects required an average of 8-10 minutes to perform the ratings task once. As the incubation literature shows, this is ample time for incubation effects to appear. Since many of the solvers in Experiment 1 reached solution during the second ratings task, it seems likely that incubation played a role in at least some of the subjects' problem solving processes.

A second concern is that the concepts themselves often served as effective clues to the problem's solution, as indicated by the number of solvers who could tell the experimenter exactly which pairs of concepts led them directly to the solution. Fortunately, this source of clues was not as helpful as might have been expected because many subjects did not recognize the value of the concepts until late in the solving session. As the protocol analyses demonstrated, those subjects who did refer to the concepts in the first solving session were more likely to solve the problem than were subjects who did not refer to the concepts until the second solving session.

Finally, the act of judging relationships in the ratings task may have helped the subjects make connections among the concepts that they otherwise might not have made. Overall, the Pathfinder ratings task may simply be too intrusive a tool for modeling problem solving in progress.
The results of the analyses of AHA! ratings for the practice problems raises some interesting questions concerning the role of the AHA! experience in the insight process. If the subjective experience of insight, the AHA! reaction, is in part dependent upon the difficulty of the problem, then the experience known as insight may not be as different from the non-insight problem solving process as is currently assumed.

The discussion of the protocol data indicated that the shotgun attracted most subjects for a great deal of their solving time. This is not too surprising since the shotgun is a highly salient and unexpected part of the problem. However, the ability to examine aspects of the problem other than the shotgun was necessary to solving the problem. Only a relatively few subjects were able to do that. There was some evidence that training helped subjects focus on the water issue as opposed to the shotgun issue. For subjects in the Passive training condition, at least, focusing on the water led to solution more often than did focusing on the shotgun.

Experiment 2

In the second experiment, the benefit of insight problem solving practice on later insight problem solving was compared to that of training a more general problem solving skill, assumption checking. Assumption checking may be a relevant skill for solving the Bar problem since there are at least two very important, but mistaken assumptions made in reading the problem. According to the protocols from the first study, most
subjects assumed that the man wanted water because he was thirsty, and that the bartender pulled out the gun for a belligerent, non-helpful purpose. The ability to recognize these assumptions and to correct them may be helpful in being able to solve the Bar problem.

In this study, four test problems were used. Three of the practice problems in the first study that were categorized as assumption-based problems were used as test problems. The assumption checking skill should be relevant for the solution of these problems. The fourth test problem in the second study was the Bar problem from Experiment 1. The Bar problem is not considered to be an assumption-based problem because the solution does not involve the correction of a single, mistaken assumption. However, assumptions do play an important role in this problem. The relative effectiveness of these two types of training (assumption checking skills and general problem solving practice) for these two types of insight problems was examined.

**Method**

**Subjects**

The subjects in this study were undergraduate psychology students at Rice University and were randomly assigned to three conditions. Subjects who did not complete the Bar problem section of the study (see Procedure section) were included in the analyses involving the three assumption-based test problems but were omitted from analyses involving the Bar problem. A total of 73 subjects were tested: 25 were assigned to the Assumption
training condition, 27 to the Practice problems condition, and 21 to the Control/No training condition. Of these subjects, 18 from the Assumption training condition, 17 from the Practice problems condition, and 16 subjects from the Control condition were included in the Bar problem analyses. Although originally a larger sample was intended for this study, the time constraints on testing subjects and the fact that many members of the subject pool were becoming familiar with the Bar problem, and thus had to be excluded from that aspect of this study, prevented the testing of a larger sample.

All subjects received course credit for voluntary participation in this research. Approximately equal numbers of males and females were randomly assigned to each condition.

**Materials**

**Control condition:** The materials for this condition were the same as in Experiment 1.

**Insight problem training:** Eight of the problems from the Active training condition in Experiment 1 were used. Two new problems were added to make a total of ten problems in the practice set. The new set of ten problems is listed in Appendix D. As in Experiment 1, five problems were assumption-based problems and five were not. While subjects worked through the ten practice problems, they also rated the problems for the level of AHA! that they may have experienced when they either solved the problem correctly or learned the solution after failing to solve the
problem. The rating scale used was a 5-point scale anchored at 1=No AHA! to 5=A Lot of AHA!.

**Assumption training:** In this condition, subjects performed a paper-and-pencil task which presented three scenarios. For each scenario, the subject was asked to generate as many common assumptions that could be made about the scenario as possible. One common assumption was given as a guide with each scenario. After each scenario was completed, a list of common assumption was provided to give some feedback to the subjects. Subjects were instructed that the skill of identifying common assumptions was important to insight problem solving and that this task was practice for a later test with insight problems. The assumption training task is presented in Appendix E.

**Test problems:** One test problem was the Bar problem from Experiment 1. The remaining three problems were earlier categorized as assumption-based problems (in Experiment 1). These three problems are listed in Appendix F.

**Procedure**

Subjects were tested individually. After signing the informed consent form, subjects were told that the experimenter was interested in training the ability to solve insight problems. Subjects were then randomly assigned to one of three conditions and the training phase was begun.

**Control:** This condition was the same as in Experiment 1.

**Practice problem training:** The experimenter explained to the subject that he or she was going to work on a set of ten insight problems as
practice for the test phase where he or she would try to solve four test problems. She then explained the AHA! ratings that the subject would make while working on the practice problems. The subject was allowed to work through the practice problems at his or her own pace, asking questions as needed.

Assumption training: The experimenter explained that insight problems are often difficult because the solver makes incorrect initial assumptions about the problem. The assumption task was explained as practice in identifying those initial assumptions. The subject was told that the assumption task is presented as training for the later insight problems. The first page of the assumption task explained again, in detail, the role of assumption checking in insight problem solving. The subject then worked through the assumption task at his or her own pace, asking questions as needed.

Test session: After the training phase was completed, the subject was shown the Bar problem and asked if it was familiar. If the subject was familiar with the problem, then that subject was not included in the part of the analysis that examined solution rates for the Bar problem or the Pathfinder analyses. The subject did, however, attempt to solve the three assumption-based problems. After completing a think-aloud practice problem (Experiment 1), the subject then tried to solve the three problems, while thinking aloud. The presentation order for the three problems was randomized for each subject. The subject was given four minutes for each problem. The subject was allowed to ask yes/no questions as in Experiment 1. After completing the three assumption-based problems, the subject then
rated each problem for the AHA! experience. The subject was then
debriefed and thanked.

If the subject was not familiar with the Bar problem, the subject then
completed the first concepts ratings task. This task was the same as in
Experiment 1 except that the instructions explicitly told the subject to make
these relatedness ratings within the context of the Bar problem. After
completing the ratings, the subject was given practice in thinking aloud and
began a ten-minute solving session with the Bar problem. The subject was
instructed to ask yes/no questions as needed. At two-minute intervals
during the solving session, the experimenter asked the subject "What are
you trying to do right now?" The subject then gave a brief statement about
the subject's current problem solving goal. At the end of 10 minutes or
when the subject solved the problem, the subject completed the second
concept ratings. If the subject did not solve the problem, he or she was not
told the solution until after completing the second ratings task. The subject
then tried to solve the three assumption-based test problems while thinking
aloud. If the each problem was not solved in four minutes, the subject was
told the answer. After finishing all four problems, the subject rated each
problem for AHA!.

Whether the subject worked the Bar problem or the assumption-
based problems first was counterbalanced across subjects. The three
assumption-based problems were always presented together and in a
random order. The time taken to solve each of the four problems was
recorded.
Results

Analysis of Practice Problems

As in the first study, subjects in the Practice condition were given ten insight problems to try to solve. There was no time limit for solution. If practice does facilitate insight problem solving, then these subjects should be more likely to solve problems at the end of the practice session than at the beginning. The number of problems solved in the first three ordinal positions was compared to the number of problems solved in the last three ordinal positions. This result was not significant. However, when solution rates for the first four problems and the last were compared, the difference was marginally significant, F(1,25)=3.71, p=.066. Subjects tended to solve a greater proportion of the last four problems (M=.65) than of the first four (M=.54). The effect of practice on solving insight problems is not as evident in this study as it was in Experiment 1. One possible explanation is that in this study, subjects were allowed to "give up" on a problem at their own discretion whereas subjects in the first study were forced to spend at least two minutes trying to solve each problem. It may be that subjects in this study quit working on some problems prematurely and so did not have as much improvement in their solution attempts as did subjects in Experiment 1.

Analysis of Insight Ratings

As in Experiment 1, subjects in the Practice Condition rated the amount of AHA! they experienced when learning the answers to the
practice problems. A one-way ANOVA compared these insight ratings for solved problems and unsolved problems. The result was similar to that in the first study. Problems that were solved were given lower insight ratings ($M=2.42$) than were problems that were not solved ($M=2.87$), $F(1,212)=10.50$, $p<0.02$.

The number of Practice Condition subjects who solved each problem was counted. The problems were then ranked in order of difficulty as determined by the proportion of subjects who solved each problem. The ten problems were divided into two groups by a median split: the five problems solved by the most subjects were labeled "Easy" and the five problems solved by the fewest subjects were labeled "Hard". Insight ratings were compared in a $2$ (Problem Type: Assumption-based vs. Non-assumption based) X $2$ (Difficulty: Easy vs. Hard) ANOVA. Neither main effects nor the interaction was significant.

One reason for this failure may be that only two problems (#115 and #202) were solved by less than 50% of the subjects in the Practice Condition; the remaining eight were solved by at least 56% to as many as 83% of the subjects. Thus, there were not that many truly difficult problems. The problems were recoded with only numbers 115 and 202 coded as "Hard" and the remaining eight problems coded as "Easy". Again, a $2$ (Problem Type: Assumption-based vs. Non-assumption based) X $2$ (Difficulty: Easy vs. Hard) ANOVA was conducted, but neither the main effects nor the interaction was significant. Thus, the relationship between problem difficulty and the subjective experience of insight remains
unclear. This analysis did not replicate the Experiment 1 result that the more difficult the problem, the higher the insight rating.

Finally, the insight ratings for the Bar problem were compared for Bar solvers and nonsolvers. The result of the one-way ANOVA was not significant. Subjects who solved the Bar problem gave an average insight rating of 4.41, while nonsolvers gave a mean rating of 4.12; the difference was not reliable. However, the mean insight ratings for the Bar problem are much higher than those ratings for the practice problems (M for Solved problems=2.42, M for Nonsolved problems=2.87). Again, a ceiling effect for the insight ratings may be in effect with the Bar problem.

Analysis of the Assumption Test Problems

The number of solvers and nonsolvers for each assumption test problem in each condition is shown in Table 9. Three Chi-square analyses were conducted on the distribution of solvers and nonsolvers across conditions (one for each problem), and none produced significant results. Subjects in all conditions were equally likely to solve each of the three assumption test problems.
Another way to look at subjects' performance on the Assumption test problems was to count the proportion of the three problems each subject solved and the time it took to solve each problem. The proportions of Assumption test problems solved were submitted to a 1-way ANOVA comparing performance across the three training conditions. The analysis did not produce a significant main effect; F(2, 70) = 0.11, p > 0.05.

Time to solution was recorded for each subject on each problem and submitted to three 1-way ANOVAs (one for each problem), testing time to solution across the three training conditions. Again, there was no difference among conditions in time to solution for each problem. Time to solution was averaged for each subject across the three test problems and submitted to a 1-way ANOVA comparing performance across the three training conditions. This analysis also failed to produce a significant effect of training condition: F(2, 70) = 0.54, p > 0.05.
Assumption-checking training did not facilitate the solving of assumption-based insight problems. This failure may be due to the nature of the three test problems. First of all, the three problems were relatively easy to solve: the average proportion solved of the three test problems was 0.66. Thus a ceiling effect may have obscured any effect of training. With more difficult problems, a training effect would be more likely to show. Secondly, three problems may not have been sufficient to demonstrate a possible training effect. Perhaps a more appropriate test of this hypothesis would require a larger set of test problems. It has been shown that a relatively large set of test problems reveals a practice effect for insight problem solving better than a small set does (Jacobs & Dominowski, 1981).

Analysis of the Bar Problem

Solution Rates: The first solution rate analysis concerned the number of subjects in each condition that solved the Bar problem at any of three points during the experiment: during the first Pathfinder ratings task, during the 10-minute solving session, or during the second Pathfinder ratings task. Table 10 shows this distribution of solvers and nonsolvers.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Nonsolvers</th>
<th>Solved at 1st Ratings</th>
<th>Solved at Session</th>
<th>Solved at 2nd Ratings</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption</td>
<td>11</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Practice</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 10
The Distribution of Bar Problem Solvers and Nonsolvers in Experiment 2
The difference in the distribution of solvers and nonsolvers across conditions was not reliable, \( \chi^2 (2, N=51) = 0.75, p >0.05 \). Subjects in each condition were equally likely to solve the problem. Unlike in the first experiment, the number of solvers in each condition was not in the predicted direction: subjects receiving assumption-checking training or simple practice in solving insight problems were not more likely to solve the Bar problem than were subjects who received no training.

The number of subjects in each training condition who solved the problem during a ratings task (either first or second) was compared to the number of subjects who solved the problem during the problem solving session; the difference was also not reliable: \( \chi^2 (2, N=22) = 1.72, p >0.05 \). Therefore, unlike in Experiment 1, there was no systematic difference among the three conditions in the likelihood of solving the bar problem during a ratings session or during the problem solving session. However, across all training conditions, more subjects solved the problem during the ratings tasks \( (n=19) \) than during the problem solving session \( (n=3) \), \( \chi^2 (1, N=21) = 324.5, p < 0.001 \).

One of the differences between the solution task in the second study and that of the first study is that subjects had less time overall to solve the problem in Experiment 2. Thus, Experiment 2 may not have provided subjects with enough time to reach solution during the problem solving. It is interesting to note, however, that a greater percentage of solvers in Experiment 2 reached solution during the first ratings task \( (14/22) \) than did solvers in Experiment 1 \( (3/20) \). Though difficult to draw any conclusions
from this disparity, it raises some questions about the efficacy of the various training tasks in the two studies. Apparently, subjects were better able in Experiment 2 to immediately solve the problem than they were in Experiment 1. However, Experiment 1 subjects were better able to reach solution during the problem solving session than were subjects in Experiment 2. It begins to appear that the solving process for subjects who can solve the problem during the ratings task is different than the process for subjects solving the problem during the solving session. From the purely subjective point of view, the process that reaches solution during the first ratings task seems to be more insightful than the other process since these subjects solved the problem without the benefit of asking questions, gathering information, or even having much time to think about the problem.

The process leading to immediate solution may, in fact, be more insightful, but it does not provide much data for study. These subjects did not provide think-aloud data, ask questions, or give goal statements. Since the majority of solvers in Experiment 2 are solvers of this type, the protocol data for solvers is scanty. Much of the remaining analysis may not be able to adequately compare solvers and nonsolvers because there are only eight solvers who provided protocol data (three solvers during the problem session and five solvers during the second ratings task).

**Question Rates:** The number of questions each subject asked was counted and divided by the amount of time spent on the Bar problem (nonsolvers=10 minutes). Subjects who solved the problem during the first ratings session (R1 solvers) asked one question (posing the solution) and
were scored with five seconds to ask it. Thus, R1 solvers were assigned a question rate of 12 questions per minute. Question rates were submitted to a 3 (Training Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solvers vs. Nonsolvers) between-subjects ANOVA. Only the main effect of Status was significant: F(1,35)=11.05, p<0.003. Solvers asked questions at the average rate of 5.49 questions per minute while nonsolvers asked questions at the average rate of 2.26 questions per minute.

However, the vast majority of solvers in this study were R1 solvers (14/22). The somewhat arbitrary assignment of a question rate of 12 questions per minute may be misleading. Being able to pose a fairly certain hypothesis, such as the solution to the problem, in five seconds does not necessarily correlate with being able to ask many questions at the rate of 12 per minute. Since Experiment 1 included only 3 R1 solvers, this dilemma was not particularly relevant; however, in Experiment 2, the preponderance of R1 solvers does pose a problem. Therefore, the question rates were re-analyzed with the R1 solvers omitted from the analysis. In this ANOVA, the effect of Status is not significant: F(1,30)=1.03, p >0.05. Bar problem solvers (n=8) asked questions at the average rate of 2.65 questions per minute, and nonsolvers (n=28) asked questions at the rate of 2.26 questions per minute. This result does not replicate the findings of Experiment 1, perhaps because of insufficient power. Removing the R1 solvers from the data drastically reduces the number of solvers from 22 to eight.

**Question Types:** Question type analyses were performed on protocol data from a subset of subjects. The eight solvers who provided protocol
data were included in the subset. From the remaining nonsolvers, four subjects in each condition were randomly chosen for inclusion. The final subset contained two Assumption training solvers, four Practice training solvers, two Control solvers, and four nonsolvers from each training condition. With so few subjects in each cell, no interactions between training condition and status (solver vs. nonsolver) were expected.

The questions asked during the problem solving session were coded into three categories: Strategy questions, Information-gathering questions, and Hypothesis-testing questions. Recall from Experiment 1 that Strategy questions asked about ways of solving the problem, the relevance of certain aspects of the problems, and indications of how close a subject was to the solution. Information-gathering and Hypothesis-testing questions were further coded into sub-categories. Information-gathering questions sought additional information about the glass of water (Water), the shotgun (Shotgun), the man's thanking of the bartender (TY), the man's departure from the bar (Leaving), or anything else (Other). Hypothesis-testing questions posed potential answers to the issues of why the man requested water (Water), why the bartender pulled out a shotgun and pointed it at the man (Shotgun), or why the man said "Thank You" (TY). In addition, Hypothesis-testing questions could also involve scenario-like answers to the problem as a whole (Global).

In the first analysis, a proportion for each question type for each subject (in the subset) was computed by dividing the number of each type of question asked by the total number of questions asked and then multiplying by 100. There were three such proportions for each subject.
These proportions were submitted to a 3 (Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solver vs. Nonsolver) X 3 (Type: Hypothesis-testing vs. Information-gathering vs. Strategy) mixed ANOVA. Question Type was the repeated measure. The main effect of question type was significant: $F(2,28)=18.66$, $p<0.001$. Information-gathering questions were asked proportionately more often ($M=53.87\%$), Hypothesis-testing questions were asked second most often ($M=37.21\%$), and Strategy questions were asked least often ($M=8.85\%$).

The Question Type X Condition interaction was also significant: $F(4,28)=3.24$, $p<0.03$. The mean proportions of each question type for each condition are shown in Table 11 and displayed in Figure 13. In general, it appears that the Assumption-training subjects are asking mostly Information-gathering questions. This interaction was further explored by analyzing the data for only one condition at a time. The proportions of each question type were submitted to three 3 (Type: Hypothesis-testing vs. Information-gathering vs. Strategy) X 2 (Status: Solver vs. Nonsolver) mixed ANOVAs, one for each training condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Assumption</th>
<th>Practice</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>2.68 %</td>
<td>5.63</td>
<td>19.31</td>
</tr>
<tr>
<td>Hypothesis-testing</td>
<td>25.40</td>
<td>52.41</td>
<td>28.75</td>
</tr>
<tr>
<td>Information-gathering</td>
<td>71.66</td>
<td>41.97</td>
<td>51.94</td>
</tr>
</tbody>
</table>
The relative proportions of Strategy, Information-gathering, and Hypothesis-testing questions asked in each condition in Experiment 2

The analysis for the Assumption-training condition revealed a main effect of question type: $F(2,8)=45.27$, $p<0.001$. The proportion of Strategy questions asked by these subjects was the smallest proportion ($M=2.68\%$) and the largest was the proportion of Information-gathering questions ($M=71.66\%$). In the middle was the proportion for Hypothesis-testing questions ($M=25.39\%$). All three means were significantly different from each other according a comparison of the means.

For the Practice condition subjects, the main effect of question type was also significant: $F(2,12)=9.38$, $p<0.004$. Again, the proportion of Strategy questions asked by these subjects the smallest proportion
The proportion of Information-gathering questions ($M=5.63\%$) and the proportion for Hypothesis-testing questions ($M=41.97\%$) and the proportion for Hypothesis-testing questions ($M=52.40\%$) were not reliably different from each other but were both significantly greater than the Strategy mean proportion. The analysis for Control subjects revealed no significant effects.

To better compare the relative proportions of each question type across training conditions, an additional analysis was performed. The proportions of Strategy, Information-gathering, and Hypothesis-testing questions asked by each subject in the subset were submitted to three $3$ (Condition: Assumption vs. Practice vs. Control) $X$ $2$ (Status: Solver vs. Nonsolver) between-subjects ANOVAs, one for each question type.

Unlike in Experiment 1, there were no differences between solvers and nonsolvers or across the training conditions in the proportion of Strategy questions asked. However, there was a significant main effect of Condition for the proportion of Hypothesis-testing questions asked, $F(2,14)=4.78$, $p<0.03$. Specifically, subjects in the Practice condition asked proportionately more Hypothesis-testing questions ($M=52.41\%$) than did Assumption-training subjects ($M=25.39\%$) or Control subjects ($M=28.75\%$). For the Information-gathering questions analysis, only the main effect of Condition approached significance, $F(2,14)=3.17$, $p=0.074$. Assumption-training subjects tended to ask proportionately more Information-gathering questions ($M=71.66\%$) than did the Practice condition subjects ($M=41.96\%$) or the Control subjects ($M=51.94\%$).

The Hypothesis-testing questions were further analyzed by subcategories. The proportion of Hypothesis-testing questions which fell into
each sub-category (Water, Gun, TY, and Solution) were submitted to separate 3 (Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solver vs. Nonsolver) between-subjects ANOVAs. No effects in any analysis were reliable. In other words, there were no differences among the training conditions or between solvers and nonsolvers in the proportion of questions asked about why the man wanted water, why the bartender pulled out a shotgun, or why the man said thank you.

A similar analysis was performed for the different categories of Information-gathering questions (Water, Shotgun, TY, Leaving, and Other). There was a significant main effect of training condition only for the proportion of Information-gathering questions in the Other category, F(2,13)=3.96, p<0.05. Specifically, Assumption-training subjects asked relatively fewer Other Information-gathering questions (M=40.08%) than did Practice subjects (M=67.66%) or Control subjects (M=62.79%). The Other category of Information-gathering questions is basically composed of questions that do not directly relate to the three primary components of the Bar problem (Water, Shotgun, Thank you).

Finally, the overall proportions of questions pertaining to the glass of water and the shotgun were computed [(Information-gathering + Hypothesis-testing)/Total Questions] and submitted to two 3 (Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solver vs. Nonsolver) between-subjects ANOVAs, one each for water and shotgun questions. No effects for the proportions of shotgun-related questions were significant. However, for the water-related questions, the main effect of Condition was significant; F(2,14)=4.28, p<0.04. Subjects in the Assumption training
condition asked proportionately more water-related questions overall ($M=23.78\%$) than did subjects in the Practice condition ($M=10.53\%$) or in the Control condition ($M=13.58\%$).

**Pathfinder Analyses—Evidence for Problem Reconstruction**

As in Experiment 1, the concept relatedness ratings subjects provided in the two ratings tasks were submitted to the Pathfinder algorithm and corresponding Pathfinder networks were produced for each subject. However, many of the solvers solved the problem during one of the ratings tasks; this can contaminate the solver Pathfinder data. Therefore, the solvers were grouped into three categories: Solvers with usable data from both Time 1 (pre-solution) and Time 2 (post-solution) ratings ($n=6$), solvers with usable data from only Time 1 ratings ($n=4$), and solvers with usable data from only Time 2 ratings ($n=7$).

Similarity scores for Pathfinder networks and correlations between concept ratings' matrices are two indicators of problem knowledge representation change. These scores were computed for the six solvers with usable Time 1 and Time 2 ratings data and for the 28 nonsolvers with usable data from both the Time 1 and the Time 2 ratings tasks. A 1-way ANOVA compared the network similarity scores for solvers and nonsolvers, but the effect of status (solver vs. nonsolver) was nonsignificant: $F(1,33)=2.15$, $p>0.05$. However, as in Experiment 1, the difference in similarity scores was in the predicted direction: Solvers' Time 1 and 2 networks were less similar ($M=0.29$), indicating greater
change in problem knowledge organization, than were nonsolvers' Time 1 and Time 2 networks (M=0.39).

The concept ratings correlations for these subjects were also submitted to a 1-way ANOVA comparing the correlations between Time 1 ratings and Time 2 ratings. Again, the effect of Status was not significant, F(1,33)=0.008, p >0.05. Solvers' mean correlation between Time 1 and Time 2 ratings was 0.62; nonsolvers' mean correlation was 0.63. Thus, these Pathfinder data do not provide overt evidence of change occurring in problem representation during the event of insight.

Network similarity scores and ratings correlations are global measures of knowledge organization using Pathfinder data. Additional analyses can be made at a more local level, that of specific links. Two links that were identified in Experiment 1 as potentially relevant to the solution of the Bar problem are the Water-Remedy and Remedy-Surprise links (Dayton et al., 1990). Chi-square analyses were performed on the number of solvers (n=15) and nonsolvers (n=28) who had each link present in either the Time 1 network or the Time 2 network. The only significant result concerned the presence of the link Remedy-Surprise in Time 2 networks: χ² (1, N=43) =10.38, p<0.002. Twelve of the 15 solvers (with usable Time 2 Pathfinder data) had this link present after solving the Bar problem compared to only eight of the 28 nonsolvers who showed this link after failing to solve the problem.

However, this result does not necessarily indicate that the link was created by the event of insight or by solving the problem. It may be that those subjects who solved the problem had the link present at Time also.
Unfortunately, there are only six solvers with good Time 1 and Time 2 data. Of these six subjects, four had this link present in their Time 1 networks. All six had this link at Time 2. Thus, most of the solvers with Time 1 data show this link in their initial representations of the problem. For the nonsolvers with good Time 1 and Time 2 data, 15 of the 28 subjects never had this link; four showed the link only at Time 2; five showed the link only at time 1; and four had the link present in both networks. The overall distribution of the Remedy-Surprise link is shown in Table 12.

<table>
<thead>
<tr>
<th>Status</th>
<th>Present at Time 1 &amp; 2</th>
<th>Present at Time 1 only</th>
<th>Present at Time 2 only</th>
<th>Absent at Time 1 &amp; 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solver</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Nonsolver</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>15</td>
<td>28</td>
</tr>
</tbody>
</table>

Another way of looking at this data is to determine how the presence of this link is related to solving the problem; in other words, how many subjects who had this link at Time 1 solved the problem compared to how many who lacked this link at Time 1 later solved the problem. Of the 13 subjects who showed the link in their initial networks, four later solved the
problem. Of the remaining nine subjects, five of them had lost the link by Time 2. Of the 21 subjects who did not have this link at Time 1, only two later solved the problem, though six others had gained the link by Time 2.

The data seem to indicate that the presence of the link between Remedy and Surprise may be important in being able to solve the problem, but the presence of the link does not seem to be caused by the event of insight itself. Thus, the primary difference in Time 2 networks between solvers and nonsolvers cannot be called a result of insight or of problem reconstruction; it appears to be more of a cause, or at least a forerunner of solution.

Pathfinder Analyses—Evidence of Training Effects

The previous analyses concentrated on the differences in networks between solvers and nonsolvers. The following analyses will examine those differences in networks across the three training conditions. As conducted in Experiment 1, ratings matrices for the concept ratings at Time 1 and Time 2 for each subject were constructed. Matrix correlations between all pairs of subjects were then computed. If the training conditions were effective in molding a subject's initial approach to the Bar problem, then ratings' correlations for subjects within the same training condition should be higher than the ratings' correlations for subjects from different training conditions.

For the analysis of Time 1 inter- and intra-condition ratings correlations, only ratings for subjects with usable Time 1 data were analyzed. Mean correlations were computed for the resulting six groups
(Assumption-Assumption, Assumption-Practice, Assumption-Control, Practice-Practice, Practice-Control, and Control-Control). Table 13 shows these correlations. Unlike Experiment 1, the inter- and intra-correlations for the training groups (Assumption and Practice) are not higher than those involving the control group.

Table 13
Time 1 Ratings' Correlations Among Training Conditions in Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Assumption</th>
<th>Practice</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption</td>
<td>0.3338</td>
<td>0.3551</td>
<td>0.3937</td>
</tr>
<tr>
<td>Practice</td>
<td></td>
<td>0.3655</td>
<td>0.4210</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td>0.4949</td>
</tr>
</tbody>
</table>

Time 2 ratings were treated similarly, using only data from subjects with usable Time 2 data. These mean correlations are shown in Table 14. Again, the inter- and intra-correlations for all the conditions are about the same. Thus, there is not much evidence in these analyses to indicate that the subjects in the training conditions began working on the problem with an interpretation different from or more consistent than that of the control subjects.
Table 14
Time 2 Ratings' Correlations Among Training Conditions in Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Assumption</th>
<th>Practice</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption</td>
<td>0.4842</td>
<td>0.4996</td>
<td>0.4899</td>
</tr>
<tr>
<td>Practice</td>
<td>0.5168</td>
<td>0.5226</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td>0.5334</td>
</tr>
</tbody>
</table>

Two further analyses examined differences in knowledge representation for solvers and nonsolvers within each training condition. These analyses only included subjects with both usable Time 1 and usable Time 2 data. A similarity score for each subject's Time 1 and Time 2 Pathfinder networks was calculated and submitted to a 3 (Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solvers vs. Nonsolvers) between-subjects ANOVA. Although the similarity scores tended to be higher for nonsolvers than for solvers (indicating less change from Time 1 to Time 2), neither Condition, Status, nor their interaction produced significant results. Network similarity scores did not differentiate solvers and nonsolvers among the training conditions. Table 15 lists the similarity scores for the three conditions.
Table 15
Similarity Scores Between Time 1 and Time 2 Networks for Solvers and Nonsolvers in the Three Training Conditions in Experiment 2

<table>
<thead>
<tr>
<th>Status</th>
<th>Assumption</th>
<th>Practice</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solvers</td>
<td>0.299</td>
<td>0.308</td>
<td>0.246</td>
</tr>
<tr>
<td>Nonsolvers</td>
<td>0.390</td>
<td>0.412</td>
<td>0.379</td>
</tr>
</tbody>
</table>

A similar ANOVA, using the correlation between each subject's Time 1 and Time 2 concept ratings was performed. Again, no effects were significant: Ratings correlations did not differ among the training conditions. Table 16 lists these ratings correlations. It is important to note that in these analyses, there are only six solvers (two from each training condition) with usable data compared to 28 nonsolvers with usable data (eleven Assumption, seven Practice, ten Control). The power of these analyses may not be strong enough to detect small but significant differences between network similarity scores or between ratings correlations.
Table 16
Correlations Between Time 1 and Time 2 Ratings for Solvers and Nonsolvers in the Three Training Conditions in Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Assumption</th>
<th>Practice</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solvers</td>
<td>0.656</td>
<td>0.613</td>
<td>0.601</td>
</tr>
<tr>
<td>Nonsolvers</td>
<td>0.584</td>
<td>0.713</td>
<td>0.597</td>
</tr>
</tbody>
</table>

Analysis of Goal Statements

At two-minute intervals during the problem solving session (four times), subjects were asked "What are you trying to do right now?" in an attempt to elicit statements about each subject's current problem solving goals. Goal statements from each subject were then coded into seven categories. Sometimes more than one goal was mentioned during a single response. Thus, some subjects could be credited with more than four goal statements. Subjects whose protocol data was examined in the question analyses were included in this analysis (one subject from this set provided no goal statements since she solved the problem in less than two minutes, and so was omitted from this analysis). The greatest number of goal statements from a single subject in this sample was six. The categories of goal statements are described in Table 17.
Table 17
Goal Statement Categories in Experiment 2

<table>
<thead>
<tr>
<th>Goal Label</th>
<th>Description/Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Determine why the man wanted water.</td>
</tr>
<tr>
<td>Shotgun</td>
<td>Determine why the bartender pulled out a shotgun and/or pointed it at the man.</td>
</tr>
<tr>
<td>Thank You</td>
<td>Determine why the man said &quot;Thank You&quot;.</td>
</tr>
<tr>
<td>Leaving</td>
<td>Determine why the man left the bar (without his water).</td>
</tr>
<tr>
<td>Strategy</td>
<td>Determine the role of the concepts from the ratings task or look for double meanings and plays on words.</td>
</tr>
<tr>
<td>Meta-Cognitive</td>
<td>Statements such as &quot;I don't know, I'm just going around in circles.&quot;</td>
</tr>
<tr>
<td>Other</td>
<td>Everything else.</td>
</tr>
</tbody>
</table>

Once all the goal statements were coded, the number of unique goal categories was counted for each subject. If a subject is fixated on a component of the problem, then he or she is less likely to produce different kinds of goal statements throughout the course of the problem solving session. On the other hand, the presence of several categories of goal statements would indicate that the subject was exploring different aspects of the problem. Statements in the Meta-cognitive category, however, do not actually represent active exploration of the problem. Fortunately, out of 81 goal statements examined, only two were placed in this category.
The number of unique goal categories for each subject was divided by the total number of goals that subject gave and then multiplied by 100. Since subjects who solve the problem before the end of the problem solving session do not necessarily have four opportunities to provide goal statements, this proportion is a more accurate measure of goal stability than the raw number of unique statements would be.

The proportions were submitted to a 3 (Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solvers vs. Nonsolvers) between-subjects ANOVA. Only the main effect of Status was significant: F(1,13)=8.14 p<0.02. Specifically, the proportion of unique goals for solvers was 77.14% compared to only 54.58% for nonsolvers. As expected, there was more change among the goal statements given by subjects who solved the problem than by nonsolvers.

To look at these statements more closely, the frequency of occurrence for each goal category was counted for the first four goal statements given (three of the 19 subjects gave only three goal statements). Table 18 lists these frequencies for the first four goal statements. Separate Chi-square analyses were performed on the frequency of occurrence for goals in the Water, the Shotgun, the Thank You, the Leaving, and the Other categories. The four Chi-square analyses were all significant: 1st Goal-$\chi^2$ (3, N=16) =40.05, p<0.001; 2nd Goal-$\chi^2$ (3, N=13) =14.13, p<0.003; 3rd Goal-$\chi^2$ (4, N=15) =16.05, p<0.03; and 4th Goal-$\chi^2$ (4, N=13) =11.38, p<0.03. The goal category appearing most often, at each goal statement, was the Shotgun category. Again, we see that the shotgun
component of the problem strongly attracts the subject's attention, and is rarely abandoned in favor of other problem components.

Table 18
The Frequency of Occurrence of Five Goal Categories for the First Four Goal Statements in Experiment 2

<table>
<thead>
<tr>
<th>Category</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Shotgun</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Thank You</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Leaving</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total # of Goals</strong></td>
<td><strong>16</strong></td>
<td><strong>13</strong></td>
<td><strong>15</strong></td>
<td><strong>13</strong></td>
</tr>
</tbody>
</table>

Since the majority of the statements given fall into the Water, Shotgun, and Thank You categories, the frequencies of occurrence of goals in these categories were compared across each condition and for solvers and nonsolvers. Only the first four goals were counted. However, in these analyses, no effect was significant. The frequencies of the Water, Shotgun, and Thank You goal categories did not differ across conditions or between solvers and nonsolvers for any of the first four goal statements.

Finally, the total proportion of goal statements belonging in the Water, Shotgun, and Thank You categories were submitted to three-way ANOVA (Condition: Assumption vs. Practice vs. Control) X 2 (Status: Solvers vs. Nonsolvers).
Nonsolvers) between-subjects ANOVAs, one for each goal category. However, no effects were significant in any analysis.

**Experiment 2 Discussion**

Experiment 2 did not show as strong an effect of practice on insight problem solving as did Experiment 1. However, it did replicate the Experiment 1 finding that the unsolved problems provided a greater subjective AHA! experience than did solved problems.

The Pathfinder data in Experiment 2 do not provide evidence of overt problem representation reconstruction. Since subjects were given explicit instructions on providing the relatedness ratings within the context of the problem, the Pathfinder data should model problem representation better in Experiment 2 than in Experiment 1. However, this representation does not change significantly when the problem is solved. Only a single link differentiates the networks of solvers and nonsolvers.

The question analyses do not provide as strong support for effects of training as was seen in Experiment 1, perhaps in part due to insufficient power. However, it may also be due, in part, to a mismatch of training and insight problem solving skills. It appears that assumption-checking training was not as helpful as initially predicted, perhaps because learning a single skill has little effect in a task requiring multiple problem solving skills. What significant differences there were in question types among the three training conditions are somewhat ambiguous. While Practice subjects tested more hypotheses, Assumption-training subjects asked more water-
questions, shown in Experiment 1 to better facilitate solution of the problem. Overall, there are simply no sufficient trends in types of questions that would indicate a strong, facilitating effect of the forms of training used in this experiment.

Finally, the goal statement analyses provided indications of what hinders problem solution. These data indicated that, in general, subjects spent far more time and resources on issues concerning the shotgun than on aspects of the problem that would be more helpful in solving the problem. The goal statements also demonstrate that some subjects were better able to shift their focus and vary their goals, indicating that they were not as "stuck" on the shotgun issues as were other subjects. Subjects who formulated several goals were more likely to solve the problem than were subjects who were less flexible.

General Discussion

The data from these two experiments present a varied, and somewhat confusing, picture of insight problem solving. First, the practice problem data in Experiment 1 show that a practice effect does exist for this type of insight problem (verbal, riddle-type problems). But the forms of training did not increase the likelihood of solving the test problem(s) in either study. Second, we see that the training did influence the rate at which subjects asked questions. In addition, the data show that subjects can be unduly influenced by the most salient aspect of an insight problem (e.g., the shotgun) to the extent that they ignore other, potentially more relevant
problem aspects (e.g., the water). However, this tendency can be reduced by some forms of training. Specifically, prior practice in solving insight problems (Experiment 1: Active training) increased the number of strategy questions asked by subjects. The value of such questions lies in the fact that they help identify the relevant and irrelevant aspects of the problem. Even the study of insight problems (Experiment 1: Passive training) increased the likelihood that subjects would examine the important problem component involving the glass of water. Finally, training in a specific problem solving skill (Experiment 2: Assumption-checking training) also facilitated subjects' tendencies to examine the role of the glass of water.

The data collected here have raised some new questions about the insight problem solving process and about insight itself. Some researchers have discussed the AHA! effect as a consequence of problem representation change: We experience this phenomenon when we recognize the new meanings of things in a new problem representation (Lockhart et al., 1988; Ellen, 1982; Kaplan & Simon, 1990). The data from the AHA! ratings in the two studies presented here indicate that this experience is also due in part to the difficulty level of the insight problem. This AHA! experience can occur even when the insight problem has not been solved. In fact, the reported AHA!s for the Passive nonsolvers in Experiment 1 were rated higher than were those for subjects who actually solved the problems. It seems that the AHA! event is not dependent upon insight, or that insight is not dependent upon solving the problem. Since the AHA! phenomenon has been used to identify insight problems and to define insight itself (i.e.:
Kaplan & Simon, 1990), this relationship between insight and the AHA! event must be better understood.

Defining insight problems is a particular problem in the insight literature. If the AHA! experience cannot be used as a defining feature, then what else differentiates insight problems from non-insight problems? Metcalfe and Wiebe (1987) discuss insight problems in terms of non-incremental, non-analytical properties of the solution process. They showed that even subjective "feelings-of-warmth" about the nearness of solution follow a non-incremental pattern during the insight problem solving process. Instead of gradually increasing as the subject neared the solution, feelings of warmth showed a step-like pattern, increasing from zero only immediately before solution occurred. Only this definition of insight problems has empirical support, but it requires that solution be attempted before a problem can be defined as an insight problem.

Weisberg and Alba (1981) describe insight problems as a set of diverse problems that are similar in that they are solved when "the subject can achieve insight into the structure of the problem" (p. 169). Some have simply identified insight problems as problems that have been previously used as insight problems in the insight literature (i.e.: Lockhart et al., 1988). These definitions have some degree of circularity to them, and so are not preferable as definitions. Finally, others have defined insight problems as those that require a change in problem representation in order to be solved (Wicker et al., 1978). This definition, too, has its limitations.

The data presented in the current studies have further complicated the use of this definition. Unlike the data presented by Dayton et al.
(1990), these studies have not provided strong support for the hypothesis that subjects who solve the insight problem have an organization of problem-related knowledge that differs from that of subjects who did not solve the problem. What differences that did exist between the solvers' and the nonsolvers' networks was small enough that we must question whether or not such a seemingly small difference constitutes true problem representation change. A change in a single link is not very strong evidence that solvers are changing their perception of the problem at the moment of insight, especially if most of the solvers started out with different links than nonsolvers. In fact, the presence of any differences in Time 2 networks for solvers and nonsolvers would not provide such evidence. A confound exists in this type of data that is not easily ferreted out. The knowledge gained with the solution of the problem could account for much of the differences between solvers' and nonsolvers' final perception of the problem. This confound between changes in problem perception due to insight and those due to learning the solution to the problem continues to complicate the issue of defining insight as problem representation change.

Given that several methodological difficulties occurred with using the Pathfinder algorithm to measure concept organization during insight problem solving, it is not surprising that such little difference between solvers' and nonsolvers' networks occurred. It is highly likely that the actual differences between solvers' and nonsolvers' problem representations could not have been captured by such a modeling technique. Consider the link Water-Remedy. This link was equally likely to be
present both before and after problem solution. However, the meaning of
this link may have changed during solution. Prior to solution, water could
be considered a remedy for thirst, while after solution, it is thought of as a
remedy for hiccups. Thus, the link could easily be present at any time, but
its meaning may not be the same across time. This kind of change cannot
be captured by the Pathfinder model as used here. While Pathfinder can
accurately model semantic distances, it is not as apt at modeling the specific
meanings of those distances. According to Ellen (1982), it is precisely the
meanings of problem-related concepts that changes during insight, not
necessarily their organization.

Whether it is their organization or their meaning that changes during
the insight process, it is still unclear how that change comes about. We
know that fixation can prevent such change (Kaplan & Davidson, 1988).
Lockhart et al. (1988) discuss the difficulty of insight problems in terms of
a "...failure to access available information" (p. 36). But what facilitates
change? The data presented here indicate that it is the recognition of
relevant aspects of the problem that allows such change to occur. Subjects
in both of the present studies focused their problem solving attempts on the
issue of the shotgun while, for the most part, ignoring other problem
components. Only when they turned their attention to the other
components of the problem and gathered information about them was
solution likely.

This "change" needed for successful insight problem solving has been
discussed in various terms: for example, changing problem spaces (Kaplan
& Simon, 1990), overcoming assumptions (Wicker et al., 1978; Dayton et
al., 1990), using appropriate conceptions (Lockhart et al., 1988), and using appropriate mental models (Montgomery, 1988). One difficulty with these definitions is that any definition of insight must be broad enough, general enough to apply to a wide variety of insight problems. Not all insight problems involve mistaken assumptions, nor do all problems initiate inaccurate conceptions of terms. While the term "problem space" is more encompassing, it is also somewhat vague. Wicker et al. (1978) tried to train a similar, broad skill which they named problem-reformulation. However, this skill was vaguely defined and based on specific processes such as identifying mistaken assumptions. Though they found some evidence of insight facilitation, they did not test this facilitation with many insight problem types. The studies here present a new candidate for this "change": the recognition and active processing of all problem components in order to find the relevant ones.

Perception of a problem includes the identification of relevant and irrelevant aspects of the problem. Usually, this perception is guided by the problem itself (Kaplan & Simon, 1990), by the solver's past experience and knowledge (Weisberg & Alba, 1982), and by the solver's prior expectations of the type of problem being dealt with (Hayes, Waterman, & Robinson, 1977). Given these various influences and constraints, it is not surprising that solvers have a difficult time identifying the relevant aspects of problems that are trying to hide those very aspects. Consider the Lake problem used as a test problem in Experiment 2:
A woman walked for twenty minutes on the surface of a lake without sinking into the water. She was not using any form of flotation device, such as a boat or a raft. How did she do it?

Granted that people attempting to solve this problem generally begin with the assumption that the water is in liquid form and that this mistaken assumption must be corrected in order to solve the problem, it is also true that people rarely begin solution attempts by questioning the condition of the lake. The problem wording focuses the solver's attention on the flotation devices, which imply a liquid medium. Thus, the initial solution attempts usually look for some method of walking on liquid water without a flotation device. However, once attention is turned toward the condition of the lake itself, solution becomes much easier, almost immediate. In the Bar problem, the salience of the shotgun's role in the problem appears to demand the solver's attention, and the role of the glass of water, another relevant component, is ignored. Problem representation change may, thus, be facilitated when all aspects of the problem are processed, rather than just a few.

The data in the two studies presented here support this hypothesis. First, most nonsolvers of the Bar problem were focused almost exclusively on the role of the shotgun, whereas solvers also explored the role of the glass of water. The goals data in Experiment 2 showed that subjects who formed many problem-solving goals, addressing different aspects of the problem, were more likely to solve the problem than were subjects who formed fewer unique goals. The data also indicate that this strategy may be facilitated by some kind of training. Research on the training of insight has
generally shown practice to be an effective form of training. Since there is a practice effect for insight problem solving, the solvers must be learning something during the practice, and that something must be general enough to apply to several types of insight problems. One possibility is that they are learning that whatever aspect of the problem initially seems most important is often not. Thus, they learn that they must consider all aspects, even ones that initially appear to be irrelevant. Training in a specific problem solving skill in Experiment 2 did not facilitate the solution of problems that utilize that skill. However, practice in insight problem solving in Experiment 1 did increase subjects' exploration of the role of the water in the Bar problem. For subjects in the Passive training condition, this exploration led more often to solution.

Another result from Experiment 1 that supports this hypothesis is the effect that the Pathfinder ratings task had on problem solution. Some subjects in that experiment (Passive training subjects) were more likely to solve the problem during a ratings task than during the problem solving sessions. One effect of the ratings task is that it forced the subjects to consider concepts (aspects) of the problem that they otherwise might have ignored.

Thus, it can be argued that here is a general problem solving strategy that may underlie the successful solution of insight problems. Even Kaplan and Simon (1990) hypothesize that good insight solvers are also "good noticers" (p. 396). Though they refer specifically to the noticing of problem invariants, problem invariants are simply aspects of a problem.
Rather than speak of problem representation change then, we may need to speak of complete problem representations.

How, then, do such complete representations come to be when not influenced by training? Kaplan and Simon (1990) propose that problem spaces are not changed or abandoned until the subject feels that the current space has been exhausted, and he or she has developed a certain level of frustration with it. At that point, he or she begins the search for a new problem space. Such a process may have occurred in Experiment 1. Recall that subjects rarely made use of the Pathfinder ratings concepts in the first solving session, yet almost immediately begin exploring those concepts at the beginning of the second session. Though incubation effects may be occurring here, it is also possible that the subject was unlikely to approach those concepts until he or she had "run out of other ideas", so to speak. In other words, during the first solving session, the more obvious hypotheses were tested. When these failed to produce the solution, the subject then turned to the next source of potential information. In this context, the ratings concepts served as an information source of problem components but were not initially recognized and used as such.

The details of this process are still unclear. Additionally, no insight study to date has examined insight across the many different types of insight problems. An adequate definition of the insight process must apply to more than simple riddle forms of insight problems as were studied here. Future research is needed to explore further the details of the process of identifying relevant problem components and the training of such a skill.
References


Appendix A

Pathfinder Demonstration

To illustrate how Pathfinder builds a network, the following set of pairwise ratings for five concepts is used. These ratings range from 1 to 5, where larger values indicate more distant relationships. This network will be constructed using the parameters \( r=\infty \) and \( q=4 \).

<table>
<thead>
<tr>
<th></th>
<th>BIRD</th>
<th>ROBIN</th>
<th>EAGLE</th>
<th>PENGUIN</th>
<th>SPARROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIRD</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>ROBIN</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>EAGLE</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>PENGUIN</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>SPARROW</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

(1=closer relationship; 5=distant relationship)

Pathfinder begins with the shortest distance estimates. BIRD and ROBIN are not already connected by any path and have the shortest distance estimate (1), so Pathfinder puts the first link between these concepts. Concepts pairs with the next shortest distance estimate (2) are then considered: BIRD-EAGLE, BIRD-SPARROW, and ROBIN-SPARROW. Because no existing path connects any of these pairs, Pathfinder places direct links here. The network now looks like this:
Next to be considered are EAGLE-ROBIN (3) and EAGLE-SPARROW (3). However, the path EAGLE-BIRD-ROBIN has a maximum path link of 2 (the maximum link weight is used when \( r = \infty \) as opposed to the sum of all the link weights in a path or the Euclidean distance between concepts). Since this path therefore has a shorter distance than the direct estimate of 3, no direct link is added between EAGLE and ROBIN. The same argument applies to EAGLE-SPARROW. The path EAGLE-BIRD-SPARROW has a shorter distance than the direct estimate, and so no link is added between EAGLE and SPARROW.

Following this line of reasoning, links are added between PENGUIN-BIRD and between PENGUIN-EAGLE, because no shorter path between these concepts exist. However, PENGUIN-ROBIN and PENGUIN-SPARROW are not directly linked. Their distance estimates are both 5, but a path through BIRD with a maximum weight of less than 5 already exists, so no additional links are added. If the distance estimate for two concepts is equal to the shortest path length connecting them, Pathfinder still adds the link. When \( q = \text{Nodes}-1 \), then paths of any length (number of links) can be considered when determining whether to add a link or not. When \( q \) is set to a smaller number, then paths with more links than that number are not considered. The final network looks like this:
Appendix B

Training problems for Experiment 1

1. A man who lived in a small town married twenty different women of the same town. All are still living and he never divorced a single one of them. Yet, he broke no law. Can you explain?
A: He is the minister. (Bowden, 1985)

2. A young boy turned off the light in his bedroom and managed to get into bed before the room was dark. If the bed is ten feet from the light switch and he used no wires, strings, or other contraptions to turn off the light, how did he do it?
A: It was still daylight. (Bowden, 1985)

3. If you have black socks and brown socks in your drawer mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure you have a pair the same color?
A: 3. Two of the three socks must be the same color. (Metcalf and Wiebe, 1987)

4. A lady got into a New York cab. She talked too much that she annoyed the cabdriver. He said, "Sorry, lady, but I can't hear a word you're saying. I'm deaf as a post and my hearing aid has not worked all day." She stopped yakking. But after she left the cab, she suddenly realized that the cabbie had lied to her. How did she know?
A: Because he took her to the right destination. (Metcalf and Wiebe, 1987)

5. Two fathers and two sons went duck hunting. Each shot a different duck, but they only shot three ducks in all. How is that possible?
A: There are only three men: a grandfather, his son, and his grandson. (Ross, Ryan, and Tenpenny, 1989)

6. Water lilies double in area every 24 hours. At the beginning of summer, there is only one lily on the lake. It takes 60 days for the lake to become completely covered with lilies. On which day is the lake half covered?
A: On the 59th day. (Sternberg and Davidson, 1982)

7. A wine bottle is half filled and corked. How can you drink all the wine without breaking the bottle or removing the cork from the bottle?
A: Push the cork into the bottle. (Gardner, 1978)

8. Sally comes to school one day, and for Show and Tell, announces that today is the birthday of both her father and her grandfather. What is more, both are exactly the same age. Her teacher tells her that this is impossible, but Sally insists that she is right. How can she be right?
A: It is her father and her maternal grandfather. (Grosswirth, Salny, and Mensa, 1981)
9. A young explorer called his expedition chief in great excitement to report that he had just found a golden coin marked 6 BC. The expedition chief immediately fired him. Why?
A: No coins made before the birth of Christ would be marked BC, because they did not know about the birth of Christ yet. (Grosswirth, Salny, and Mensa, 1981)

10. A man moors his boat in a harbor at high tide. A ladder is fastened to the boat with three rungs showing. The rungs are 12 inches apart. At low tide, the water level sinks twenty feet. How many rungs are now showing?
A: 3 rungs. As the water level sinks, so does the boat. (Grosswirth, Salny, and Mensa, 1981)
Appendix C

Concepts Used for Pathfinder Models

<table>
<thead>
<tr>
<th>Within the problem</th>
<th>Related to solution but not in problem</th>
<th>Related to problem but not to solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shotgun</td>
<td>Surprise</td>
<td>Pretzels</td>
</tr>
<tr>
<td>Thank You</td>
<td>Remedy</td>
<td>TV</td>
</tr>
<tr>
<td>Man</td>
<td>Relieved</td>
<td></td>
</tr>
<tr>
<td>Bar</td>
<td>Friendly</td>
<td></td>
</tr>
<tr>
<td>Glass of water</td>
<td>Loaded</td>
<td></td>
</tr>
<tr>
<td>Bartender</td>
<td>Paper bag</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Training Problems for Experiment 2

1. Two men played five games of checkers and each man won the same number of games. There were no ties. How could this be?
   A: They were not playing each other. (Sternberg and Davidson, 1982)

2. A young boy turned off the light in his bedroom and managed to get into bed before the room was dark. If the bed is ten feet from the light switch and he used no wires, strings, or other contraptions to turn off the light, how did he do it?
   A: It was still daylight. (Bowden, 1985)

3. If you have black socks and brown socks in your drawer mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure you have a pair the same color?
   A: 3. Two of the three socks must be the same color. (Metcalf and Wiebe, 1987)

4. A lady got into a New York cab. She talked too much that she annoyed the cabdriver. He said, "Sorry, lady, but I can't hear a word you're saying. I'm deaf as a post and my hearing aid has not worked all day." She stopped yakking. But after she left the cab, she suddenly realized that the cabbie had lied to her. How did she know?
A: Because he took her to the right destination. (Metcalf and Wiebe, 1987)

5. A police officer had a brother but the brother had no brother. How is that possible?
A: The police officer is a woman. (Ross, Ryan, and Tenpenny, 1989)

6. Water lilies double in area every 24 hours. At the beginning of summer, there is only one lily on the lake. It takes 60 days for the lake to become completely covered with lilies. On which day is the lake half covered?
A: On the 59th day. (Sternberg and Davidson, 1982)

7. A wine bottle is half filled and corked. How can you drink all the wine without breaking the bottle or removing the cork from the bottle?
A: Push the cork into the bottle. (Gardner, 1978)

8. Sally comes to school one day, and for Show and Tell, announces that today is the birthday of both her father and her grandfather. What is more, both are exactly the same age. Her teacher tells her that this is impossible, but Sally insists that she is right. How can she be right?
A: It is her father and her maternal grandfather. (Grosswirth, Salny, and Mensa, 1981)
9. A young explorer called his expedition chief in great excitement to report that he had just found a golden coin marked 6 BC. The expedition chief immediately fired him. Why?
A: No coins made before the birth of Christ would be marked BC. because they did not know about the birth of Christ yet. (Grosswirth, Salny, and Mensa, 1981)

10. A man moors his boat in a harbor at high tide. A ladder is fastened to the boat with three rungs showing. The rungs are 12 inches apart. At low tide, the water level sinks twenty feet. How many rungs are now showing?
A: 3 rungs. As the water level sinks, so does the boat. (Grosswirth, Salny, and Mensa, 1981)
Appendix E

Assumption Training Task for Experiment 2

Instructions
One of the many reasons that solving insight problems is difficult is that often the insight problem is set up in such a way that the solver makes some incorrect assumptions about the problem. These incorrect assumptions then interfere with finding the correct solution.

The following problem is a classic example of how a mistaken assumption completely prevents the solution of a problem.

A child is injured in an accident and rushed to a hospital. The doctor takes one look at the child and says: "I can't treat him, that's my son." The doctor is not the child's father. How do you explain this?

Most people (even the most liberal) initially assume that the doctor is a male. If this is so, then the problem appears impossible to solve. However, that assumption is wrong. The problem does not say that the doctor is a male. It is simply assumed. And the correct answer is that the doctor is female and is the boy's mother.

Though this problem appears trivial, there are many other problems where the incorrect assumptions are not so easily identified. In this experiment, you will be given some training or practice in identifying assumptions so that they can be checked. You
will then be able to use this ability in solving a set of somewhat more difficult insight problems.

In the following task, you will be given scenarios of fairly common events. You will only be given a few details about each event. Your job is to identify as many common assumptions that people might reasonably make about the details of each scenario. After you have finished each scenario, you can compare your assumptions with some that we came up with. There is no "correct" set of assumptions for each scenario, and there is no limit on how many assumptions can be made. This task is simply designed to give you practice in thinking about assumptions in this manner.

The first scenario is on the next page. Read the scenario and think of what picture that scenario brings to mind. Your mental picture probably has more detail to it than what the scenario actually provides. These details are assumptions. So write down these things that you think could be reasonably assumed about the scenario. These assumptions may come from stereotypes or from your own experiences. One common assumption is provided to help you start. When you are finished, turn the page to see some of our assumptions. There are 3 scenarios in all. You have about 10-15 minutes to complete them all.

Scenarios (example assumptions were presented on separate pages from the scenarios, and each scenario appeared on separate pages)
Scenario 1:

Two men enter a convenience store at 11:30 on Friday.

Assumption 1: The men are together, not separate people who don’t know each other.

Your Assumptions:

Some Scenario 1 Assumptions
It is 11:30 p.m.
The men want to buy either beer, cigarettes or fast food.
They are not grocery shopping.
They are not buying pantyhose or other typically feminine products.
They want to buy alcohol before midnight when it becomes illegal to sell alcohol.
They might be robbers.
They are at least 18 years old since they are called "men".
They are not employees of the store.

Remember: These assumptions are not necessarily true. They are possibilities that need to be verified before they can be considered facts. See how wrong these assumptions are if the scenario was actually describing an inspection of the store by the president and the vice-president of marketing for that company.
Scenario 2:

Chris calls Mrs. Johnson to ask for some money.

Assumption 1: Mrs. Johnson is a relative of Chris', probably Chris' mother.

Your Assumptions:

Some Scenario 2 Assumptions

Chris is male.
He needs money for himself.
He is broke.
He is probably between the ages of 18 and 25.
He is probably a student.
He knows Mrs. Johnson personally.

See how wrong these assumptions are if the scenario was actually as follows: Chris is a 45-year old female senior fund raiser for a local charity. She is contacting Mrs. Johnson for a donation to that charity. She has never met Mrs. Johnson before.
Scenario 3:

A woman is filling a prescription for oral contraceptives at the pharmacy.

Assumption 1: The pills are for herself.

Your Assumptions:

Some Scenario 3 Assumptions
She is sexually active or is about to be.
She wants the pills for their contraceptive properties.
She does not want to get pregnant.
She plans on ingesting the pills.
Her age is past puberty and before menopause.
She has a prescription for them.
Her prescription came from a gynecologist.
She is filling the prescription for intervals of one month.

How many of your assumptions would be correct or relevant if the scenario actually described a pharmaceutical researcher who was studying the effect of oral contraceptives on weight gain in adult females?
Appendix F

Assumption-Based Test Problems for Experiment 2

1. A man who lived in a small town married twenty different women of the same town. All are still living and he never divorced a single one of them. Yet, he broke no law. Can you explain?
A: He is the minister. (Bowden, 1985)

2. Two fathers and two sons went duck hunting. Each shot a different duck, but they only shot three ducks in all. How is that possible?
A: There are only three men: a grandfather, his son, and his grandson. (Ross, Ryan, and Tenpenny, 1989)

3. A woman walked for twenty minutes on the surface of a lake without sinking into the water. She was not using any form of flotation device, such as a boat or a raft. How did she do it?
A: The lake was frozen. (Bowden, 1985)