Memory for impasses during problem solving

ANDREA L. PATALANO and COLLEEN M. SEIFERT
University of Michigan, Ann Arbor, Michigan

Three experiments were conducted to investigate the relative memorability of solved versus unsolved problems in long-term memory. In each experiment, subjects worked on a set of potentially solvable word problems, with the time spent on each problem held constant. Problem memorability was then measured with a free-recall task. In Experiment 1, in which a majority of problems were solved, unsolved problems were better remembered. In Experiments 2 and 3, we expanded on these results by manipulating problem difficulty and thus the ratio of solved to unsolved problems. When unsolved problems were as frequent as or more frequent than solved problems, no memory differences were found. Across all three experiments, the ratio of solved to unsolved problems was found to be a significant predictor of unsolved-problem memorability, but was not significantly related to the memorability of solved problems. The results illustrate that when impasses in problem solving are infrequent, they are more available in memory than are solved problems. It is speculated that this memory phenomenon may facilitate the recognition of opportunities to return to problems that have been terminated short of solution.

In the world, we are constantly faced with problems to solve. Sometimes we are able to solve a problem at the moment that it arises; however, at other times, we reach an impasse in problem solving and are forced to terminate work on a problem before arriving at a solution, in order to pursue other goals. As problem solvers, it may be to our advantage to return to these pending problems at later points in time. Sometimes these problems simply must be satisfied eventually; but even for less critical goals, it is reasonable to assume that similar problems may be encountered again, and so reaching some resolution may be helpful to later experiences. Protocol studies of planning have shown that subjects can recognize opportunities to realize pending goals (Hayes-Roth & Hayes-Roth, 1978). For example, 1 subject noticed an unanticipated shortcut while mentally simulating an intended path for executing a series of errands. The subject incorporated an opportunity to buy a newspaper at a magazine stand on the new path into his plan, even though he had originally intended to achieve this goal elsewhere. Because circumstances in the world may have changed since a problem was first attempted, a successful solution may be more likely if problem solving is resumed at a later moment in time, under more opportune circumstances.

Given the utility of returning to and reattempting unsolved problems, is there any evidence that pending problems are more likely to be retrieved from memory? In a classic experiment, Zeigarnik (1927) provided evidence

that this ability might be due to heightened memory access to interrupted problems. She gave subjects a series of manual and mental tasks to perform, such as counting beads, creating a dog figure out of clay, and constructing cardboard boxes. The subjects were allowed to complete half of the problems, but were then interrupted by the experimenter. On a free-recall task administered after the problem-solving session, the subjects recalled more of the interrupted than the completed problems, suggesting that interruption by the experimenter made those problems more memorable than the ones that the subjects were allowed to complete. This memory effect for interrupted problems is not well substantiated, and differences such as time spent on problems and the nature of the problem-solving processes invoked may contribute to repeated failures to replicate the Zeigarnik effect (Seifert & Patalano, 1991; Van Bergen, 1968). However, a similar effect was demonstrated by Yaniv and Meyer (1987) in a word-definition task, in which subjects were presented with definitions for familiar yet infrequent words and asked to name the word. They reported that many of the words they could not recall were on "the tip of their tongue." Later, faster lexical decisions were made for the words that had been unsuccessfully retrieved, indicating heightened activation of the "tip-of-the-tongue" words. These results provide some evidence that unresolved problems may be more available in memory.

Problems in which a solution cannot be reached due to an impasse in problem solving differ from these phenomena in a number of important ways. The nature of interruption, as operationalized by Zeigarnik, referred to an arbitrary and immediate experimenter-generated stopping point imposed on a task for which a solution plan is readily apparent to the subject (e.g., how to count beads). Unlike Zeigarnik (1927), we are interested in cases in which the subjects themselves reach an *impasse* in the problem-

solving process, rather than simply being interrupted by the experimenter in the midst of an easy task. "Impasse" refers to the point at which one is "stuck" in problem solving (VanLehn, 1988), reflecting a lack of knowledge about how to proceed in solving a problem. In addition, in contrast to Yaniv and Meyer (1987), we are interested in memory for newly encountered problems, not previously learned words. Yaniv and Meyer's (1987) paradigm established heightened activation resulting from a retrieval failure for lexical entries already present in memory (e.g., Meyer & Schvaneveldt, 1971). However, for impasses in problem solving, no solution has yet been discovered or encoded into memory, so any processes affecting memory availability must go beyond facilitating retrieval of past solutions.

Several theoretical motivations suggest that impasses in problem-solving attempts (as defined above) might be more memorable than successfully solved problems. Van-Lehn (1990) has stressed the importance of impasses in procedural skill learning. Protocol evidence illustrates that procedural skill revision is motivated by reaching an impasse in applying a procedure. According to VanLehn, it is at this point that one realizes that the procedure is inadequate, and searches for a way to revise the procedure so as to get past the impasse. This kind of learning is termed failure-driven learning (Schank, 1982), because learning begins when some part of an active procedure fails. This theory posits more elaborate processing at points of impasse (see Craik & Tulving, 1975, regarding depth of processing) because one must apply potential strategies, weigh alternative approaches, and discover a solution path in order to resolve the impasse. Thus, unsolved problems might be more memorable than solved ones because, by definition, unsolved problems require more elaborate processing than problems for which a solution procedure readily comes to mind.

On the other hand, examples can be drawn from the memory literature illustrating that well-understood situations, such as solved problems, are often more memorable than ones that are poorly or incompletely understood. Pachauri (1935) found that subjects had better memory for completed tasks than for interrupted ones on some manipulations of Zeigarnik's original experiment. He attributed these results to the fact that "once finished, a task becomes in a sense an entire or fixed form ... this difference can be invoked as an explanation of the greater recall of the [completed items]" (p. 377). Bartlett (1932) illustrated that people are more likely to misremember story details that do not conform to a memory schema, while at the same time exhibiting accurate recall for details consistent with schematized knowledge. Similarly, Dooling and Lachman (1971) found better memory for lists of ambiguous sentences when the lists were given organizing titles prior to being read than for sentences that were not disambiguated with a title. Bransford and Johnson (1972) also found better memory for sentences disambiguated with a cartoon illustration than for the same sentences given without the illustration. These results suggest that solved problems might be more memorable to the extent that they are organized around complete problem-solving schemas. Unsolved problems, on the other hand, might not be memorable because important aspects of the problem (e.g., problem representation, start and goal states, available operators) might not be organized by a single schema.

Three experiments were conducted to investigate the relative availability of successful solution attempts versus impasses in problem solving in long-term memory. In particular, we addressed the following question: Is there evidence of differential memory access for solved versus unsolved problems? And, if so, what are the conditions that influence the relative memorability of each class of problems? In Experiment 1, subjects were instructed to work on a series of word problems, and problem memorability was measured with a free-recall task. In Experiments 2 and 3, we expanded on the results of Experiment 1 by manipulating problem difficulty and thus the frequency of solved and unsolved problems.

EXPERIMENT 1

In Experiment 1, we investigated the relative memorability of solved versus unsolved problems under timed trial conditions. The subjects were given a set of ordinary word problems (e.g., Mosler, 1977) with the instructions to work on each problem until they either reached a solution or "got stuck." Time was explicitly controlled by allowing the subjects only 1 min in which to work on each problem. The problems were pretested to ensure that all of them were solved by a substantial number of pilot subjects and that, when a problem was solved, the solution was usually generated in less than 1 min. This procedure made it likely that, if a subject knew how to solve a problem, it could be solved within the allotted time. However, if a solution procedure was not known, an impasse would be reached within the minute, and the time constraint would only serve to prevent the subject from further deliberation over the problem. Despite a superficial resemblance of this impasse paradigm to Zeigarnik's (1927) interruption studies, the two are not the same. Zeigarnik interrupted subjects in the midst of their active pursuit of a solution. In the present study, we attempted to ensure that subjects were simply prevented from further efforts on problems in which they had either solved the problem or had already reached an impasse on their own.

Because there may be important differences between problems that are easily solved and those that reliably result in impasses among subjects, the experiment included only problems that had been solved by a subset of pilot subjects. As a result, although each subject experienced impasses, these impasses occurred on different problems for different subjects. This ensures that any observed memory differences cannot be attributed to factors specific to particular problems. Also, "being at an impasse" and "having solved a problem" are perceivable mental states, so we wished to verify that our assignments of com-

pletion status to attempted problems were consistent with the subjects' own perceptions. Therefore, the subjects were asked to circle their solution, if they had one, in order to make clear when they believed that they had arrived at a reasonable solution.

After attempting all of the problems, the subjects were given a free-recall task to assess memory for answered versus unanswered problems (e.g., Zeigarnik, 1927). On the basis of the hypothesis that unsolved problems are more memorable than solved ones, it was expected that the subjects would recall a greater proportion of problems that they had left unsolved.

Method

Subjects. Sixty-nine undergraduates at the University of Michigan (34 female and 35 male) participated in this experiment. The subjects received credit toward an introductory psychology course.

Materials. Thirty word problems, requiring mathematical, logical, and spatial reasoning skills were used (from Friedland, 1970; Morris, 1988; Mosler, 1977; Muller, 1989; see the Appendix for examples). All of the problems were pretested on a separate group of subjects and were selected on the basis of rate of successful completion; only problems that were solved correctly within 1 min by at least a subset of pretest subjects were used. Each problem was presented to the subjects on a separate sheet of paper, with space below the problem for work and solution to be recorded. Each workbook contained a problem-solving instruction sheet, 30 pages of word problems presented in a different random order to each subject, and a final recall task sheet.

Design and Procedure. The subjects were tested in four groups of 10 to 20 in 1-h sessions. At the outset of each session, the problem-solving instructions were read aloud while the subjects followed along in their workbooks. They were told that they would be presented with a series of word problems, one on each of the subsequent pages of the workbook, and that they would be given exactly 1 min in which to work on each problem. They were instructed to work consistently and diligently throughout the experiment, making every effort to solve each problem, and to record all work and any solutions in the space provided. If they arrived at a solution before the minute had elapsed, they were instructed to spend the remaining time checking their work. At the end of the 1-min trial, the subjects were instructed to stop and circle their answer, if they had completed one, and then to proceed to the next trial. This procedure continued until all 30 problems were attempted.

Immediately following the last problem, the subjects were given a free-recall task. The experimenter read the instructions aloud; specifically, they were asked to recall as many problems as they could from the first part of the experiment, jotting down problems on the sheet provided in the order in which they came to mind. They were instructed to write only enough information so that someone else could recognize the problem to which they were referring. After 3 min had elapsed, the subjects were asked to stop working on the task.

Results and Discussion

Problem-solving results. Prior to examining the recall data, all the problems were scored by the experimenter as being solved or unsolved on the basis of the information written by the subjects. All circled responses were scored as solved, because the subjects had been instructed to circle only their problem solutions; the rest were scored as unsolved. On average, more problems were solved (20.0 problems; 67%) than left unsolved (10.0 problems;

33%) by each subject. Solved problems were further scored as either correct or incorrect in their solutions. A problem was scored as correct if the subject's response matched the answer on a prepared answer sheet (e.g., if the subject wrote "1 cup" for the first problem in the Appendix). It was scored as incorrect if the subject wrote a plausible but incorrect response (e.g., if the subject wrote "2 cups"). All the circled responses fit one of these two categories. More problems were solved correctly (11.1 problems; 56% of solved problems) than incorrectly (8.9 problems; 44% of solved problems).

Free-recall results. Free-recall data were scored by counting a response as an instance of recall whenever it uniquely identified one of the test problems. More specifically, the experimenter, blind to problem solution status, compared each recall response with a master list of test problems. The subjects tended to describe problems by their superficial features; for example, one problem (Problem 1 in the Appendix) might be recalled as "figuring out how much coffee you drank," whereas another (Problem 2 in the Appendix) might be labeled "the one with the different-colored paint cans." Because each problem was unique in its superficial features, matching responses to test problems was straightforward, and all responses were successfully scored using this criterion. Overall, a mean of 10.8 problems (36%) were recalled. Significantly more of the unsolved problems (45%; $SD = .24$) than of the solved problems (33%; $SD = .16$) were recalled [$t(68) = 3.64, p = .001$]. Among solved problems only, problems solved correctly were more likely to be recalled (37%; $SD = .20$) than those solved incorrectly [28%; $SD = .21$; $t(68) = 2.57, p = .012$]. However, unsolved problems were also more likely to be recalled than correctly solved problems [$t(68) = 2.44, p = .017$].

EXPERIMENT 2

Although the procedure employed in Experiment 1 was successful in creating a situation in which subjects solved some problems and reached impasses on others, the generality of the conclusion that unsolved problems are more memorable than solved ones is limited by the fact that the subjects solved many more problems than they left unsolved. Specifically, solved problems occurred twice as frequently as unsolved problems when averaged by subject. This difference is potentially important because the relative size of sets from which items are to be recalled has been shown to affect free recall through output interference (Crowder, 1976). Other research (Hastie & Kumar, 1979) also suggests that people often have better memories for items on a list that constitute the smaller of two identifiable subsets of a list. Additionally, the Von Restorff (1933) effect predicts that isolated items will be more memorable than related items within a series. In Experiment 1, unsolved problems might have been perceived as isolated events, simply because they were more

sparsely dispersed among solved problems for most of the subjects. Thus, it might be that the subjects recalled more unsolved problems in Experiment 1 because unsolved problems constituted a smaller, more unusual set.

In Experiments 2 and 3, we investigated this hypothesis by repeating the procedure of Experiment 1, but using problems of greater difficulty and a shorter solution time interval. The intent in Experiment 2 was to have subjects successfully solve fewer problems and reach impasses more frequently than in Experiment 1, thus eliminating any set-size differences. The goal of Experiment 3 was to increase problem difficulty even further, resulting in less frequent occurrence of solved problems. If it is the case that unsolved problems are more memorable than solved ones, regardless of relative set sizes, then unsolved problems should be more memorable in both Experiments 2 and 3. If it is not unsolved problems per se, but the smaller set that is more memorable, then there should be no memorability differences in Experiment 2, and solved problems should be more memorable in Experiment 3. Finally, if there is an interaction between solution status and memorability such that unsolved problems are more memorable but only when they are infrequent, then we should see no differences between solved- and unsolved-problem memorability in either of the following experiments.

The procedure used in Experiment 2 was similar to Experiment 1; however, an alternative manipulation check was performed on each trial. The subjects rated the perceived completion status of each problem on a scale from 0 (*I did not arrive at a solution to this problem*) to 4 (*I am extremely confident that my solution to this problem is the correct one*). Our intention was, again, to assess subjects' notions of problem status by using a more sensitive measure than the circling criterion employed in Experiment 1.

Method

Subjects. Seventy-one undergraduates at the University of Michigan (34 female and 37 male) participated in this experiment. The subjects received credit toward an introductory psychology course.

Materials. Thirty word problems, requiring mathematical, logical, and spatial reasoning skills were used in this study (from Friedland, 1970; Morris, 1988; Mosler, 1977; Muller, 1989). All of the problems were pretested on a separate group of pilot subjects, who were given 45 sec in which to work on each problem. Problems were then selected on the basis of rate of successful completion (answered by more than 1 but fewer than half of the subjects).

Problem workbooks were set up that were identical to those used in Experiment 1, with the exception that an additional rating scale appeared at the bottom of each problem page. Rather than circling their solutions, like the Experiment 1 subjects, the subjects in Experiment 2 were asked to rate how confident they were that each answer was correct. Rating scales ran from 0 to 4, with 0 corresponding to *I did not arrive at a solution to the above problem*, and 1-4 corresponding to increasing degrees of confidence in a solution, from *I am extremely doubtful that my solution is the correct one* to *I am extremely confident that my solution is the correct one*.

Design and Procedure. The subjects were tested in three groups (n = 10, 11, and 50, respectively) in 1-h sessions. At the outset of each session, the problem-solving instructions were read aloud while the subjects followed along in their workbooks. They were

given the same instructions as in Experiment 1, with two exceptions: they were told that they would have 45 sec rather than 1 min to work on each problem, and they were instructed to circle a rating of each problem at the end of the 45-sec period rather than circle the solution, as in Experiment 1. The remainder of the procedure was identical to that of Experiment 1.

Results and Discussion

Problem-solving results. Prior to examining the recall data, all problems were scored on the basis of solution information written by the subjects. An independent coder, given an answer sheet, scored each problem as either solved or unsolved. A problem was scored as unsolved if the space below the problem was either left blank, or if no part of the work on the page met the solution constraints of the problem. For example, if the solution was supposed to be a certain number of cups (as in Problem 1 in the Appendix), and an incomplete equation (e.g., "1 − 1/6 + ...") was written, the problem was scored as unsolved. A problem was scored as solved if the form of the response met the solution constraints of the problem. For example, "2 cups" would be scored as a solution to Problem 1. All responses fit into one of the two categories. On average, slightly fewer problems were solved (13.5 problems; 45%) than left unsolved (16.5 problems; 55%) by each subject. In addition, each solved problem was scored as either correct or incorrect, by using the same criteria as in Experiment 1. All responses again fit one of the two categories. Slightly fewer problems were solved correctly (6.1 problems; 45% of solved problems) than incorrectly (7.4 problems; 55% of solved problems). Compared with Experiment 1, in which solved and unsolved problems occurred in a 2:1 ratio on average, set-size differences were considerably reduced in this experiment.

Free-recall results. Free-recall data were scored as in Experiment 1. All responses were successfully scored using this criterion. Overall, a mean of 11.2 problems (37%) were recalled. Approximately the same percentages of unsolved problems (36%; SD = .17) and solved problems (40%; SD = .20) were recalled [$t(70)$ = −1.50, p = .139]. Thus, when we eliminate set-size differences between solved and unsolved problems, we also eliminate memorability differences. Among solved problems, the recall rate was higher for problems solved correctly (47%; SD = .25) than for those solved incorrectly [33%; SD = .25; $t(70)$ = −3.77, p < .001], as in Experiment 1. Unlike Experiment 1, however, correctly solved problems were more memorable here than were unsolved problems [$t(70)$ = −3.71, p < .001].

Rating results. Twenty subjects who failed to use intermediate values on the rating scale were excluded from this analysis. Among the remaining subjects, a mean of 53% of all problems were rated as "unsolved," and the remaining 47% were rated as solved. A breakdown by individual ratings is as follows: A confidence rating of 0 (*I did not arrive at a solution to the above problem*) was assigned to a mean of 53% of the problems, low-confidence ratings of 1 and 2 (*I am extremely/somewhat*

doubtful that my solution is the correct one) were each
assigned to 10% of the problems, and high-confidence
ratings of 3 and 4 (*I am somewhat/extremely confident
that my solution is the correct one*) were assigned to 13%
and 14% of the problems, respectively.

Overall, a high correlation was found between confi-
dence category (0, low, and high) and experimenter-judged
solution status [χ^2 (4, N = 2,130) = 1,896.56, p <
.00001]. This illustrates that the experimenter's coding
of subject responses was consistent with the subjects' own
intuitions about their performance. A one-way between-
subjects ANOVA was used to compare the recall rates
for problems in each rating category: 0 = 35% (*SD* =
.18), 1 = 33% (*SD* = .28), 2 = 34% (*SD* = .34), 3 =
34% (*SD* = .29), and 4 = 53% (*SD* = .34). A signifi-
cant effect of rating was found [$F(4,47)$ = 4.84, p =
.002]. Problems rated 4 were found to be significantly
more memorable than problems rated 0–3 [$F(1,50)$ =
18.75, p < .001], which did not differ significantly from
each other, consistent with the finding that correctly solved
problems were more memorable than incorrectly solved
or unsolved problems.

The results of Experiment 2 suggest that the hypothesis—
that impasses alone result in greater problem memorability—
is false. Otherwise, unsolved problems would have been
recalled more often than solved ones, as in Experiment 1.
In Experiment 2, however, successful-solution trials and
impasse trials occurred about as often (45%–55%); no
differences in memory availability based on solved or un-
solved completion status were found. Instead, recall data
based on unsolved, correctly solved, and incorrectly solved
categories showed that problems that were correctly solved
were recalled more frequently than others. This enhance-
ment of recall for correct problems is supported by the
ratings analysis, which showed that problems rated as
most confidently correct were the best recalled. Thus,
when solved and unsolved problems occur approximately
equally often, no memory advantage for unsolved prob-
lems is observed.

EXPERIMENT 3

In Experiment 1, enhanced memorability for unsolved
versus solved problems was observed under conditions
in which impasses occurred more frequently. In Experi-
ment 2, in which the frequency of solved and unsolved
problems was nearly equal, no memory differences for
solved versus unsolved problems were found. At the least,
these results make clear that impasse status alone does
not account for greater problem memorability. In fact,
it may be that relative set size, rather than impasses per se,
predicts problem memorability if problems from a smaller
set are always more memorable. However, it is still pos-
sible that an interaction occurs between set size and solu-
tion status, such that unsolved problems are indeed more
memorable than solved ones, but only when the unsolved
problems occur less frequently. Other studies suggesting
set-size effects (Hastie & Kumar, 1979) as an explana-
tion for item memorability showed that people had better

memories for incongruent items on a list that happened
to constitute the smaller of two identifiable subsets of the
list. However, these studies did not consider what would
happen if incongruent behaviors made up the larger set,
and offered no conclusions about the relative roles of set
size and solution status (or degree of congruency) in pre-
dicting item memorability.

To answer this question, the roles of completion status
and set size in the memorability of solved versus unsolved
problems must be examined in cases in which a majority
of problems are unsolved. In Experiment 3, we inves-
tigated these factors by repeating the procedure used in
Experiment 2, except using problems of even greater dif-
ficulty. In this way, subjects may solve fewer problems
than those on which they reached impasses, achieving the
opposite of the frequencies in Experiment 1. If it is the
case that memory is better for the problems in the smaller
set, then the subjects in this experiment should show bet-
ter memory for solved problems. However, if it is the
case that memory is better for infrequent, unsolved prob-
lems, then this manipulation should lead to the same re-
sults as in Experiment 2.

Method
Subjects. Sixty-one undergraduates at the University of Michi-
gan (31 female and 30 male) participated in this experiment. Fifty-
one received credit toward an introductory psychology course; 10
were paid for their participation.

Materials. Thirty word problems, requiring mathematical, log-
ical, and spatial reasoning skills, were used in this study (from Em-
met, 1968; Friedland, 1970; Morris, 1988; Mosler, 1977; Muller,
1989). To increase problem difficulty, many were modifications
of Experiment 2 problems. All problems were pretested on a sepa-
rate group of subjects to ensure that, although fewer subjects solved
each problem, each problem was still solved by some of the pilot
subjects within the 1-min time limit. In the actual study, problem
workbooks were set up, identical to those used in Experiment 1.

Design and Procedure. The subjects were tested in groups of
5–15 in 1-h sessions. At the outset of each session, the problem-
solving instructions were read aloud while the subjects followed
along in their workbooks. The subjects were given the same in-
structions as in Experiment 1, with the exception that they were
no longer instructed to circle problem solutions.

In accordance with the instructions, the subjects were given 1 min
in which to work on each problem. At the end of each 1-min pe-
riod, they were asked to go on to the next problem. This procedure
continued until all 30 problems were completed. The recall-task
instructions and procedure that followed were identical to those used
in Experiment 1.

Results and Discussion
Problem-solving results. As in Experiment 2, all prob-
lems were scored as solved or unsolved on the basis of
solution information written by the subjects. On average,
fewer problems were solved (11.98 problems; 40%) than
were left unsolved (18.02 problems; 60%) by each sub-
ject. The ratio of solved to unsolved problems was 2:3,
almost the reverse of Experiment 1. Solved problems
were further scored as either correct or incorrect in their
solutions, as in Experiment 2. More problems were solved
correctly (6.61 problems; 56% of solved problems) than
incorrectly (5.38 problems; 44% of solved problems).

Free-recall results. Free-recall data were scored as in Experiment 2. All responses were successfully scored using this criterion. Overall, a mean of 8.0 (27%) problems were recalled. This recall rate is somewhat lower than in the prior two experiments, suggesting that poorer overall memory may result when a large proportion of the trials results in unsolved problems. For solution status differences, approximately the same percentages of unsolved problems (28%; $SD = .13$) and solved problems (25%; $SD = .15$) were recalled [$t(60) = 1.25, p = .216$]. Among solved problems only, there was no difference in recall for problems solved correctly (26%; $SD = .20$) and those solved incorrectly [24%; $SD = .22$; $t(60) = .56, p = .580$]. The results illustrate that when the set-size ratio is reversed so that solved problems are in the majority, no memorability differences based on solution status are evident. These results suggest that it is not set size per se, but an interaction between set size and completion status that leads to increased problem memorability. This hypothesis will be investigated further by analyses combining the observations from all three experiments.

Combined Analyses of Experiments 1, 2, and 3

Although solved versus unsolved problems had a ratio of 2:1 in Experiment 1, nearly 1:1 in Experiment 2, and 2:3 in Experiment 3, these ratios were averages over the set of subjects within each experiment. The ratio of solved to unsolved problems for each individual varied; some of the subjects solved more than they left unsolved, while others actually solved fewer than they left unsolved. Thus, we could further examine set-size effects by looking across the three experiments at individual set sizes, which ranged from 29:1 (solved to unsolved) to 3:27. With this rationale, two alternative sets of analyses were performed on the pool of subjects from all three experiments to better examine the relative influences of completion status and frequency of solution on recall.

In the first set of analyses, only the subjects who fell into one of the following categories were considered: those who solved at least two thirds of the problems, or those who left at least two thirds of the problems unsolved. Additionally, 3 subjects who solved fewer than 3 problems were removed from the latter group so that the groups could be compared in an equal range of set sizes (these 3 subjects were removed from all further analyses). The subjects with less extreme set-size differences were not included in this analysis because of the difficulty in determining when set-size differences influence behavior. For example, how does one determine whether a ratio of 14 unsolved problems to 16 solved problems is perceived as a set-size difference or as approximately equal sets? Each group, ranging from 3 to 10 problems in its smaller set, was then analyzed separately. Fortuitously, the two groups contained approximately the same number of subjects, and the mean number of problems in each group's smaller set were roughly equivalent (7.2 unsolved problems in the first group vs. 7.8 solved problems in the sec-

ond group). Analyses performed on each of these two groups were similar to those performed in individual experiments.

Unsolved < Solved. The subjects who solved more problems than they left unsolved constituted 46 out of the 139 cases: 37 subjects from Experiment 1, 8 from Experiment 2, and 1 from Experiment 3. In this group, a mean of 7.2 problems (24%) were unsolved, and 22.9 problems (76%) were solved. Of the solved problems, 11.9 (52%) were solved correctly, and 11.0 problems (42%) were solved incorrectly. A paired t test showed a significant difference in the recall percentage of unsolved (45%; $SD = .25$) versus solved (36%; $SD = .14$) problems [$t(45) = 2.36, p = .023$]. Although correctly solved problems (41%; $SD = .18$) were better recalled, than incorrectly solved problems [32%; $SD = .19$; $t(45) = -2.63, p = .012$], their recall did not differ significantly from that for unsolved problems [$t(45) = -1.0, p = .322$]. These results are consistent with the results obtained in Experiment 1, in which more unsolved problems than solved problems were recalled, and enhanced memory for unsolved problems was observed.

Unsolved > Solved. The subjects who solved fewer problems than they left unsolved constituted 47 out of the 139 cases: 5 from Experiment 1, 22 from Experiment 2, and 20 from Experiment 3. In this group, a mean of 22.2 problems (74%) were unsolved, and 7.8 problems (26%) were solved. Of the solved problems, 4.2 (54%) were solved correctly, and 3.6 (46%) were solved incorrectly. A paired t test showed no significant difference in the recall percentage of unsolved (31%; $SD = .13$) versus solved (34%; $SD = .23$) problems [$t(46) = 1.10, p = .279$]. These results are consistent with the results obtained in Experiment 3, in which no difference in memory for solved versus unsolved problems was found when solved problems were in the larger set. Also, as in Experiment 3, a comparison between correctly solved (38%; $SD = .20$), incorrectly solved (30%; $SD = .32$), and unsolved problems showed no difference in means [$F(2,45) = 1.64, p = .206$].

Thus far, the combined results further support the hypothesis that it is the interaction between set-size and problem memorability that leads to heightened memorability. More specifically, comparing the recall means across the two groups, there is evidence that less frequent occurrence is related to unsolved-problem memorability but not to the memorability of solved problems. This is evidenced by the fact that when solved problems were in the smaller set, they were not recalled more often than when they were in the larger set (34% vs. 36%). The recall of unsolved problems, however, greatly increased when they were in the smaller set than when they were in the larger set (45% vs. 31%). To further investigate this interaction, a set of regression analyses was run on the combined data for all of the subjects (including those whose set-size ratios were between 11:19 and 19:11). To examine the effect of set size on the recall of unsolved problems, percent recall of unsolved problems was regressed onto the number of
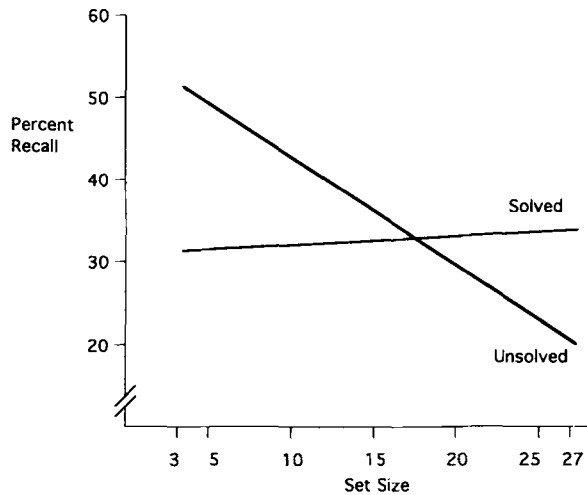
Figure 1. Percent recall of solved versus unsolved problems as a function of set size.

unsolved problems for each subject. The regression line of best fit is illustrated in Figure 1. Though set size did not account for all of the variance ($R^2 = .15$), it was a statistically significant predictor of percent recall of unsolved problems [$b = -1.3\%$; $t(197) = 5.86, p = .0001$]. However, when percent recall of solved problems was regressed onto the number of solved problems for each subject, set size had no statistically significant effect [$R^2 = .001$; $b = .2\%$; $t(197) = 1.06, p = .289$]. Additionally, the latter regression coefficient was positive, suggesting, if anything, a slight increase in recall with an increase in set size rather than the reverse. Figure 1 compares this regression line with that generated for unsolved problems.

The results of the combined analyses might best be described as follows. First, the memorability of the solved problems appears to be independent of number of solved problems in a set. Roughly the same percentage of solved problems was recalled across all set sizes. However, unsolved-problem recall appears to be related to the frequency of occurrence of unsolved problems within the sequence of trials. As the number of unsolved problems decreases, the memorability of these problems increases. As a result, when unsolved problems are infrequent in a problem set, their memorability will increase relative to solved problems, and a greater percentage of them will be recalled. When unsolved problems occur frequently, however, their memorability will not increase, and they will be recalled at the same rate as solved problems.

GENERAL DISCUSSION

Why might frequency of occurrence affect problem memorability for unsolved problems, but not for solved problems? One possible factor is the more general phenomenon of expectation failure. When events deviate from one's predictions, the resulting processing appears to re-

sult in better memory for those events, compared with those consistent with one's expectations (Schank, 1982). For example, work in schematic memory has shown that deviations from event predictions are better recalled than actions predicted by an active script in a story-recall task (Bower, Black, & Turner, 1979). Similarly, it has been found that discrimination of atypical actions is easier than discrimination of typical script actions across stories based on the same script (Graesser, Woll, Kowalski, & Smith, 1980). In more realistic task settings, Greenwald and Sakumura (1967) found that information novelty enhances the acquisition of propaganda, and Hamilton and Gifford (1976) found that illusory correlations are made between infrequent behaviors and minority groups more often than any other combination of behavior and group set sizes. Von Restorff's (1933) finding that better memory occurs for isolated items within a series, and Hastie and Kumar's (1979) result indicating better memory for incongruent behaviors (and that the relative memorability increased as their number decreased) may similarly reflect that special processes invoked when novel or unexpected events occur may result in differential encoding and enhanced memorability.

How might expectation failure account for the lack of differences in memory for solved problems, regardless of frequency? If one enters the problem-solving situation with expectations of success, and if these expectations are consistent with the majority of cases observed (as in solved trials), then there is no reason to change expectations nor to attend carefully to the particular instances experienced. This situation would be analogous to that in which Hastie and Kumar (1979) presented subjects with a description of an individual followed by a list of predominantly congruent traits, or to that in which the subjects in the present Experiments 1 and 2 solved the majority of problems they attempted. However, when one is presented with a preponderance of information that violates his or her expectations, as when expected success is met with repeated failures (unsolved problems), these failures may be noted as deviations from expectation. However, eventually, an abundance of contradictory instances may lead to a change in expectations, so that failures are now expected, and successes, when they do occur, are seen as isolated events. In any case, failures of the initial expectation may result in greater attention to trials that are both consistent and inconsistent with original expectations, leading to similar memorability at test time. This situation may reflect the experience of the subset of subjects in the present Experiments 2 and 3, who solved fewer problems than they left unsolved. Some support for success expectations comes from Rosenzweig (1943), who found better memory for correctly solved problems when subjects were told that a set of problems constituted an intelligence test and therefore expected more frequent failure instead of success on the presumably difficult problems.

Of course, there may be other plausible explanations for the results showing the interaction of solution status and frequency of occurrence. In addition, other factors

affecting problem-solving performance may be reflected in set-size differences. For example, in Experiment 3, in which more problems were left unsolved on average, poorer overall memory was observed. For these subjects, repeated trials in which the problems were left unsolved may have resulted in decreased motivation and attention to the tasks. Potentially, some of the memorability effects for both unsolved and solved problems would depend upon some minimal effort put toward a solution. Sets in which many failures are experienced may impact other factors that are important to the problem-solving process. Another potential factor is that failures that occur early in a sequence are of interest, and thus may result in enhanced memory effects, but that as more failures are experienced, less interest and attention is paid to later failures. The memory effects observed may be due to better memory for only the initial failures experienced in a problem set, followed by a fatigue effect. In any case, much more must be determined in order to ascertain what specific factors in the problem-solving process affect the memorability features observed. However, we can conclude from these studies that, under certain conditions, problem completion status may be an important predictor of problem memorability. Specifically, when a problem set includes fewer instances of impasses in problem solving, those events appear to be more easily accessible in memory.

Finally, some results indicate that correctly solved problems are sometimes more memorable than incorrectly solved problems, regardless of the relative sizes of the problem sets. Of course, correctly solved problems are more likely than either incorrect or unsolved problems to be encoded in terms of problem schemas already in memory, and problem schemas have been shown to be particularly prominent in subjects' behavior when solving word problems (VanLehn, 1989). To the extent that memory for the details of an event is improved when that event is consistent with a preexisting schema (Bartlett, 1932), there is reason to believe that correctly solved problems will be unconditionally more memorable than incorrectly solved problems, and more memorable than unsolved problems when other factors (such as set size) do not favor the memorability of unsolved problems. Rating data from Experiment 3 showed that subjects are able to discriminate among problem types, further supporting the hypothesis that correctly solved and incorrectly solved problems are processed differently.

The psychological reality of solution status as a predictor of problem memorability has significant implications for theories of problem solving. First, the evidence presented here suggests that theories appealing to problem status as an explanatory factor must more carefully consider the processing variables associated with Zeigarnik-like effects (Baddeley, 1963; Yaniv & Meyer, 1987). That is, memorability effects attributed to Zeigarnik (interruption) effects may be more accurately described in terms of reaching impasses in some solution attempts. For example, in Yaniv and Meyer's (1987) studies, only one third of the word-definition trials resulted in a failed retrieval attempt, indicating that subjects more often solved the definition than met an impasse. Under those conditions, impasse memorability rather than Zeigarnik-like interruption appears to be involved. In addition, although some theories have attempted to characterize the processing path in problem solving (Laird, Newell, & Rosenbloom, 1987; Newell & Simon, 1972; VanLehn, 1989), there has been no connection between processing characteristics and problem memorability. The present results suggest that the nature of the processing occurring on each solution attempt may affect memorability and therefore, presumably, representational differences for problems based on the processing that occurs. These results may provide an important constraint with which to assess performance of computational models of cognition (Newell, 1990). Finally, recent computer models in the area of case-based reasoning (Hammond, 1989; Kolodner, 1983; Riesbeck & Schank, 1989; Schank, 1982) have focused on the retrieval of previously solved as well as unsolved problems from memory. The memorability effects observed here may be an important indicator of how retrieval is accomplished within the cognitive system.

The present results are also consistent with intuitions about the reactivation of goals in real-world problem solving. Arguably, the experience of attempted, potentially solvable problems that result in an impasse are relatively less frequent and isolated among more easily accomplished undertakings. Because day-to-day efforts usually result in more goals accomplished than problems left as "stuck," better memory access to unsolved goals may occur as a natural consequence of goal pursuit behavior. As a result, enhanced memory for unsolved problems might usefully lead to later retrieval and, potentially, a more successful resolution of past unsolved goals. If unsolved problems are readily available in memory, one may be more likely to notice opportunities to solve these problems, and then to recall and reattempt pending unresolved problems. Thus, the present research raises the further question of when, and under what circumstances, the memory characteristics of past problems may promote the return to pending problems in memory.

REFERENCES

BADDELEY, A. D. (1963). A Zeigarnik-like effect in the recall of anagram solutions. *Quarterly Journal of Experimental Psychology*, 15, 63-64.

BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11, 177-220.

BRANSFORD, J. D., & JOHNSON, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, 11, 717-726.

CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.

CROWDER, R. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

DOOLING, D. J., & LACHMAN, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 88, 216-222.

EMMET, E. R. (1968). *Brain puzzler's delight*. New York: Emerson Books.

FRIEDLAND, A. (1970). *Puzzles in math and logic*. New York: Dover.

GRAESSER, A. C., WOLL, S. B., KOWALSKI, D. J., & SMITH, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 503-513.

GREENWALD, A. G., & SAKUMURA, J. S. (1967). Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality & Social Psychology*, **7**, 387-397.

HAMILTON, D. L., & GIFFORD, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, **12**, 392-407.

HAMMOND, K. J. (1989). *Case-based planning*. San Diego: Academic Press.

HASTIE, R., & KUMAR, A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality & Social Psychology*, **37**, 25-38.

HAYES-ROTH, B., & HAYES-ROTH, F. (1978). A cognitive model of planning. *Cognitive Science*, **3**, 275-310.

KOLODNER, J. L. (1983). *Retrieval and organizational strategies in conceptual memory*. Hillsdale, NJ: Erlbaum.

LAIRD, J. E., NEWELL, A., & ROSENBLOOM, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, **33**, 1-64.

MEYER, D. E., & SCHVANEVELDT, R. W. (1971). Facilitation and recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, **90**, 227-234.

MORRIS, S. (1988). *Omni games*. New York: Omni.

MOSLER, G. (1977). *The puzzle school*. New York: Abelard-Schuman.

MULLER, R. (1989). *The great book of math teasers*. New York: Sterling.

NEWELL, A. (1990). *Unified theories of cognition*. Cambridge: Harvard University Press.

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

PACHAURI, A. R. (1935). A study of gestalt problems in completed and interrupted tasks. *British Journal of Psychology*, **25**, 365-381, 447-457.

RIESBECK, C. K., & SCHANK, R. C. (1989). *Inside case-based reasoning*. Hillsdale, NJ: Erlbaum.

ROSENZWEIG, S. (1943). An experimental study "of repression" with special reference to need-persistence and ego-defense reactions to frustration. *Journal of Experimental Psychology*, **32**, 64-74.

SCHANK, R. (1982). *Dynamic memory*. New York: Cambridge University Press.

SEIFERT, C. M., & PATALANO, A. J. (1991). Memory for incomplete tasks: A re-examination of the Zeigarnik effect. In *Proceedings of the Thirteenth Annual Cognitive Science Society* (pp. 114-119). Hillsdale, NJ: Erlbaum.

VAN BERGEN, A. (1968). *Task interruption*. Amsterdam: North-Holland.

VANLEHN, K. (1988). Toward a theory of impasse-driven learning. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 19-41). New York: Springer-Verlag.

VANLEHN, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527-579). Cambridge, MA: MIT Press.

VANLEHN, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

VON RESTORFF, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschungen*, **18**, 299-342.

YANIV, I., & MEYER, D. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology*, **13**, 187-295.

ZEIGARNIK, B. (1927). Das Behalten erledigter und unerledigter Handlungen. *Psychologische Forschungen*, **9**, 1-85.

APPENDIX
Example Problems Used in the Experiments

1. You have a cup of coffee before you of which you drink one sixth. You then refill the cup. You next drink one third from the cup. Again you refill the cup. Finally, you drink one half of the cup. How much coffee did you drink?

2. You walk into a paint shop and the manager has mislabeled all of the paint cans. A green label means red paint, a blue label means brown paint, a red label means black paint, and a black label means green paint. What color indicates blue paint?

3. Howard, Tom, and Jack are all members of the Freedman family. One evening, at a family dinner, they were discussing their relative ages. If Tom is twice as old as Howard will be when Jack is as old as Tom is now, who is the oldest, next oldest, and youngest?

4. What is the largest sum of money in current U.S. coins (but no silver dollars) that a person can have in his pocket without being able to give someone change for a dollar, half-dollar, quarter, dime, or nickel?

5. Make the equation true by moving only one line: II = VI.

6. In a parking lot, seven company automobiles are lined up in a row in seven parking spots.
    There are two vans, both of which are adjacent to the same sports car.
    There is one station wagon that dates back to 1969.
    There are two limousines, one black and one white, which are never parked adjacent to one another.
    One of the two sports cars is always on one end.
    The parking lot is located at the south end of the building.

    If one of the sports cars is in the second spot, which spot is the station wagon in?