

The Role of Predictive Features in Retrieving Analogical Cases

HOLLYN M. JOHNSON AND COLLEEN M. SEIFERT

University of Michigan

Access to prior cases in memory is a central issue in analogical reasoning. Previous research accounts for access in terms of overall similarity between complete new exemplars compared to complete stored instances and stresses the relative importance of surface-level similarities in access to complete cases (Gentner & Landers, 1985; Rattermann & Gentner, 1987). However, for cross-domain reminders, abstract similarities capture the important commonalities between cases (Schank, 1982; Seifert, McKoon, Abelson, & Ratcliff, 1986). Therefore, models of analogy must account for structural-level reminders when they do occur in terms of abstract similarities. In planning and problem-solving tasks, a stored exemplar may be more useful if accessed *before* the new pattern is complete, when past experience can bring to bear possible solutions or warn of potential dangers while the outcome is yet undetermined. Further, different partial sets of abstract features may result in differing access to analogous cases. Features that *predict* when prior cases might be useful to problem solving could serve as better retrieval cues than other abstract cues that are equally similar, yet less distinctive to the specific problem situation. To test these hypotheses, several experiments were conducted using thematic stories in a modification of the reminding paradigm developed by Gentner and Landers (1985). By examining the relative effectiveness of subsets of features in accessing relevant cases, it was found that a subset of abstract cue features predicting when a planning failure might occur led to more reliable access to complete prior analogies than did a subset of abstract features expressing specific information about planning decisions and outcomes. Further experiments show that how distinctly the feature sets characterize the conditions leading up to the planning decision point, and not differences in the overall similarity to the case, determines access based on abstract cues. © 1992 Academic Press, Inc.

Previous research has demonstrated the utility of analogical reasoning in learning (Pirulli & Anderson, 1985; Ross, 1989b) and in problem solving (Gick & Holyoak, 1980). Novices often rely on specific examples as they learn a new task, before they learn general principles (Pirulli & Anderson, 1985), and comparing a series of analogous problems can lead to induction of more abstract problem-solving schemas (Gick & Holyoak, 1980; Ross, 1989a). Others have

proposed models of planning, such as case-based reasoning, that involve accessing cases from memory and applying them to new situations (Hammond, 1989; Kolodner, 1983; Riesbeck & Schank, 1989). However, before one can map a previous example to a current problem, one must retrieve a relevant candidate case from memory.

By most accounts, retrieval depends on how similar the new problem is to an example stored in memory. In Tversky's (1977) influential model, similarity is a function of the number of shared features minus the number of distinctive features, with each shared feature contributing equally to overall similarity. Other researchers have proposed that overall similarity determines access to potential analogs in memory (Anderson, 1986; Gentner & Landers, 1985). Much recent work, however, has focused on distinguishing different kinds of

This work was supported by subcontract of the Office of Naval Research under Contract N0014-91-J-1128 to the University of Michigan. Thanks to Douglas Medin for comments on an earlier manuscript, and to Kris Hammond and Steve Lytinen for discussions regarding the research. Eric Stone and Keith Smith provided invaluable statistical assistance. Thanks also to Candace Maisner and Michelle Berris for assistance in conducting the experiments. Reprint requests should be addressed to Colleen Seifert, University of Michigan, 330 Packard Rd., Ann Arbor, MI 48104.

similarity and determining how they may affect retrieval. A distinction has been made between structural and superficial features (Gentner, 1983; Holyoak, 1985), where the former are relations among problem elements and the latter are features that do not enter into any relational structure. Gentner and associates (Gentner & Landers, 1985; Rattermann & Gentner, 1987) found that surface or superficial features result in more frequent access, and Ross (1987) has provided further evidence that different kinds of superficial similarity (story line versus object correspondence) lead to different rates of access, although both studies found a significant number of reminders based on structural features alone. Novick (1988) has found that novices and experts attend to different kinds of similarity, and other work (Seifert, McKoon, Abelson, & Ratcliff, 1986) has found that retrieval can occur based solely on more abstract, relational features.

Some types of abstract similarities that instances may have in common have been characterized by research investigating knowledge structures in memory. Some proposed knowledge structures include plot units (Lehnert, 1980), macrostructures (van Dijk, 1979), thematic organization points (TOPs) (Schank, 1982), and thematic abstraction units (TAUs) (Dyer, 1983). They capture shared relationships between concepts, such as interactions between goals and plans (Schank, 1982), without being context dependent. The features involved in these abstract characterizations could thus be termed "structural." Such knowledge structures have received psychological validation in work by Seifert et al. (1986), Lehnert, Black, and Reiser (1981), and Seifert, Dyer, and Black (1986).

Although some research has found successful retrieval based on abstract features (Gentner & Landers, 1985; Rattermann & Gentner, 1987; Seifert et al., 1986), little work has been done on the role of different kinds of abstract similarities in accessing instances stored in memory. Similarities

based on abstract features are particularly important in accounting for the subset of reminders that are *not* based on superficial similarities. In cross-contextual reminders, abstract strategies from one domain are applied in another, such as taking the "fork" strategy from chess and developing it as the "option" play in football (Collins, 1987). When one encounters a problem in a novel domain, one may find few superficial similarities available for retrieving a relevant and potentially helpful instance from memory. In other situations, ubiquitous superficial features may not be sufficient to distinguish one prior case from others in memory. Because such cross-domain reminders do occur naturally, it is important to understand how this access based on structural similarity occurs.

The theory of case-based reasoning (Hammond, 1989; Kolodner, 1983; Riesbeck & Schank, 1989; Schank, 1982) proposes that memory access to abstract analogies is determined on the basis of the functionality of case information. In a functionally organized memory, one encodes goal-relevant cues, such as ones that predict potential problems before they occur. Such cues are used to index knowledge structures containing planning information on how to avoid those problems. With these features as indices in memory, one will retrieve previous instances having those features in the course of understanding a new case.

This functional perspective assumes it will be advantageous to retrieve a prior case based on abstract similarities at the time when they can provide information that is helpful in the current situation. This would require indexing prior cases based on the abstract features that will be apparent to the processor at the time when retrieval would be most useful. For example, in planning situations, this point would occur when one knows the planning conditions and constraints and is considering what action to take. Retrieving an appropriate case from memory at that point could

provide a potential solution, or a warning about a past planning failure to avoid. In the latter case, features that predict when a potential planning failure might occur would be the most useful retrieval cues. Therefore, if human memory is indeed functionally organized, cases sharing abstract features predictive of future planning failures should lead to better access than cases sharing other abstract but nonpredictive features. In either case, retrieval would not be based on a whole-to-whole match (as in Gentner & Landers, 1985, and Rattermann & Gentner, 1987), but on matching from a current incomplete exemplar.

Different subsets of abstract features could lead to better access for several reasons. One is that predictive feature sets may simply contain more information than other subsets of abstract features and thus are more similar to past cases in memory. However, under a functional memory hypothesis, even when two sets of cues have equivalent similarity to a target case in memory, features related to the planning decision should produce a retrieval advantage. Predictive features may more distinctly characterize relevant planning cases, leading to reminders appropriate to the planning decision and avoiding retrieval of irrelevant cases. Other types of feature sets may share features with a number of different planning situations and thus could lead to many spurious reminders in addition to appropriate ones.

The experiments presented here investigate this "predictive-features" hypothesis. However, before investigating what types of abstract features lead to useful access to cases, the utility of subsets of abstract features in retrieval must be established. Previous work on access in analogical reminding (Gentner, 1983; Gentner & Landers, 1985; Rattermann & Gentner, 1987) has used examples and stories with complete structures. In addition, most knowledge structure theories try to characterize whole events, themes, or episodes (Dyer, 1983; Schank, 1982). For example, a planning ep-

isode would include planning conditions, constraints, the action taken, and the outcome. Prior work on successful access to analogies, then, has either focused on matching entire test cases to cases in memory, or has not systematically examined the issue of incomplete feature sets (Gentner & Landers, 1985; Holyoak, 1985; Ross, 1989a). A first experiment will determine whether one can in fact reliably access old instances based on a "partial" versus a "complete" set of features. To support the predictive-features hypothesis, it must first be established that such partial feature sets provide sufficient commonalities to allow access to previous cases in memory.

INTRODUCTION TO EXPERIMENTS

All of the experiments used stories based on thematic abstraction units (TAUs) (Dyer, 1983), which were selected because of their relevance to planning. TAUs are based on abstract interactions of goals and plans as reflected in familiar cultural adages (such as, "counting your chickens before they've hatched"), especially those involving common expectation and planning failures, and so are likely to be familiar to subjects. The fact that the adages underlying TAUs have been developed within this culture suggests that they express knowledge valuable in many situations (e.g., not just those involving "chickens"). TAUs may contain planning information about problems that can occur and how to avoid or solve them and serve to organize storage of individual episodes in memory, so that cases having the same theme are stored with similar indices. For example, Table 1 shows two stories that can be characterized with the abstraction "counting your chickens before they've hatched," but which contain different content features. The thematic relations among the stories used in these experiments were validated by Seifert et al. (1986), who found that subjects could reliably detect these thematic similarities and use them to sort the stories into distinct

TABLE 1
SAMPLE STUDY AND TEST STORIES

Study story

Judy was overjoyed about the fact that she was pregnant. She looked forward to having a baby boy, and wanted one so badly she felt absolutely certain it would be male. As a result, she bought all kinds of toy cars, trucks, miniature army soldiers, and even arranged an extravagant "It's a boy" party. Finally, the big moment came, and she was rushed to the hospital. Everything went smoothly in the delivery room, and at last she knew. Judy's lively bouncing baby was actually a girl.

Complete-theme test story

Harrison disliked his small apartment and shabby furniture. His rich aunt Agatha was near death, and although he hadn't seen or spoken to her in 15 years, he felt assured of inheriting a great fortune very shortly because he was her only living relative. He had already thought of plenty of ways to spend a lot of money fixing his place up. Confident of his inheritance, Harrison began charging everything from color televisions to cars to gourmet groceries. When Aunt Agatha finally died and her will was read, she had left all her millions to the butler and now Harrison was in debt.

Predict-theme test story

Harrison disliked his small apartment and shabby furniture. His rich aunt Agatha was near death, and although he hadn't seen or spoken to her in 15 years, he felt assured of inheriting a great fortune very shortly because he was her only living relative. He had already thought of plenty of ways to spend a lot of money fixing his place up.

Outcome-theme test story

Confident of his inheritance, Harrison began charging everything from color televisions to cars to gourmet groceries. When Aunt Agatha finally died and her will was read, she had left all her millions to the butler and now Harrison was in debt.

groups, even though the stories differ in setting and other contextual features.

Experiments 1 and 2 used a reminding paradigm based on Gentner and Landers (1985), where a study set of stories are read, followed by a later test where cue stories are presented and subjects are asked to report any reminders of the study stories. Although this task does not directly involve problem solving, Ross (1987) has supported these reminding results using a problem-solving task. This suggests that the reminding paradigm allows examination of access in circumstances that do apply to problem solving. Further, truly spontaneous reminders may be rare in problem-solving situations as well. Even in research involving a problem-solving paradigm (e.g., Reiser and Faries, 1988; Ross, 1989a), subjects are encouraged to refer back to earlier examples in solving new problems, which can be viewed as a hint that old instances are relevant and should be accessed.

In the present experiments, subjects studied TAU-based stories in the first phase, and then in a later phase were tested

with stories based on the same TAUs originally studied, but sharing no surface or content features. Each TAU-based test story had three versions: a complete theme, a predict theme, and a theme outcome. Table 1 shows the test versions of a sample story and its paired study story. Experiment 1 used test stories having either a full set of thematic features (complete theme) or a subset of those features that provided only the initial thematic elements (predict theme) to compare access based on incomplete exemplars to that found with complete overlap of abstract features. To investigate the effectiveness of predictive features versus other potential sets of incomplete thematic features, Experiment 2 used as test stories both "predict-theme" stories, which presented the subset of thematic elements leading up to a decision point and "theme-outcome" stories, containing the subset of thematic features including just the planning decision and its outcome. In Experiments 3 and 4, subjects rated overall and thematic similarity, respectively, for each test story version to its

paired study story with the same TAU pattern. Using a direct comparison instead of a memory task assesses whether similarity alone could account for the rates of access in Experiment 2. Finally, Experiment 5 used the same set of predict-theme and theme-outcome stories as in Experiment 2, but had subjects perform the matching task by comparing the study and test stories directly rather than from memory, in order to further assess the two story types' distinctiveness within the set of study stories.

EXPERIMENT 1

This experiment tested whether subjects could reliably use abstract features to access cases, even when given an incomplete feature set as a cue. Subjects were tested with (a) full-length thematic stories that presented initial conditions potentially leading to a planning failure, the planning decision made, and its outcome (complete-theme), or (b) thematic stories that only set up conditions for a planning failure but did not allow certain identification of a theme (predict-theme). If the predict-theme stories, which present only a partial thematic structure, lead to reminders at a rate no higher than that expected by chance, this would cast some doubt on the predictive-features hypothesis. Further, by comparing predict-theme story results to those for complete-theme stories, we can put bounds on the potential effectiveness of partial feature sets.

Method

Subjects. Fifty-four University of Michigan undergraduates participated in a single session lasting about 1 h. They received course credit in an introductory psychology class for participating. Subjects were run in groups of two to six.

Materials. The materials consisted of eight pairs of stories, each set based on a different thematic abstraction unit (TAU), as developed by Dyer (1983), and four additional thematic stories without matches, used as fillers (see Table 1 for sample test

and study stories). All the stories reflected familiar adages such as "don't count your chickens before they've hatched." Seven story pairs were drawn from materials used by Seifert et al. (1986) in testing the validity of these knowledge structures, and an eighth pair was added. These stories included thematic, but not contextual, similarities. The four filler stories also contained adage-based themes, but they did not duplicate those used in the target story pairs.

One story from each of the eight target pairs was presented in a study packet. Each story ranged from 75 to 95 words. Six study packets were prepared, each containing the same eight stories in random order. Test packets contained the other eight stories from the target pairs and four filler stories.

Each of the eight target stories in the test packet was presented in one of two versions: a complete-theme version (80-95 words) that included the initial elements of the thematic pattern as well as the conclusion or outcome of the story and a predict-theme version (40-50 words) that duplicated the initial elements of the complete-theme version but did not include the conclusion or resolution of the story. Table 2 shows an example of the thematic fea-

TABLE 2
EXAMPLE OF PREDICTIVE FEATURES IN ONE STORY

Key: A letter by a story element means that the story version contains that element.

- S Study story
- C Complete-theme test story
- P Predict-theme test story
- O Outcome-theme test story

Theme: Counting your chickens before they've hatched

Elements:

- | | |
|-------|---|
| S C P | (1) X desires A. |
| S C P | (2) X assumes A. |
| S C P | (3) X does not act to ensure or verify A. |
| S C O | (4) X invests resources based on A. |
| S C O | (5) A does not occur. |
| C O | (6) X has used up resources for nothing. |

tures present in each type of story. Each predict-theme story provided enough information to set up conditions under which the planning failure reflected in its theme could occur, but was truncated to exclude information about whether a planning failure did occur and what its outcome was. The predictive features selected were those that one could perceive *before* the decision point in a planning situation and that would reasonably lead to expectations about what might occur next. Appendix A shows the adage-based themes and thematic features present for each target story pair used.

Each subject received four predict-theme test stories and four complete-theme test stories, each uniquely corresponding to one of the study stories. The pairing of study story to complete-theme or predict-theme test story was counterbalanced across subjects, and all 12 test stories were placed in randomized order for each subject. Each subject received the same four filler stories.

Each of the test stories was followed by these instructions taken from Gentner and Landers (1985):

IF THIS STORY REMINDS YOU OF A STORY FROM THE FIRST PART OF THE EXPERIMENT, PLEASE WRITE OUT THE MATCHING STORY AS COMPLETELY AS YOU CAN. TRY TO INCLUDE THE NAMES OF CHARACTERS, THEIR MOTIVES, AND WHAT HAPPENED.

Blank lines were provided for subjects to describe the study story brought to mind. To allow for multiple reminders, additional space was provided for each test story, along with instructions to add any other reminders.

Procedure. The reminding procedure developed by Gentner and Landers (1985) was modified to allow testing in a single, 1-h session by including a short intervening task. Subjects first received a study packet and were instructed to read through the stories and study them for 5 min, as they would be tested on them later. All subjects read all the stories at least once in this study period. After 5 min the study packets were

collected and subjects worked on an unrelated distractor task for 10 min. Then, in the test phase, subjects received the test packets and verbal instructions that they should work forward in the test booklet and not return to prior pages. They were also told that there were no answers to some of the reminding problems in the test packet, and if an answer did not come to them relatively quickly they were to go on to the next page.

After finishing this phase, free recall memory for the study stories was tested. Subjects were asked to write down brief phrases identifying stories they could recall from the study packet. They were given a few minutes to do this and were not allowed to look back through the test packet.

Results

Each page of the test packet was coded by story type (complete-theme, predict-theme, or filler) and response type (match, mismatch, or null). Mention of the corresponding story from the study packet counted as a match, whether it occurred as the first or second response on a page; only 3.3% of all responses were listed as second responses. Filler stories by definition had no matches. A mismatch was coded if the response indicated a story in the study set that did not have the same theme as the test story. A null was coded if the subject made no response or made a response that could not be identified as one of the study stories. The latter, termed intrusions, accounted for 4.6% of the responses.

Subjects on average wrote 4.4 responses as reminders in the test portion and recalled a mean of 4.8 study stories in the free recall at the end of the experiment. The overlap of these two measures was high. For each subject, two conditional probabilities were calculated: (a) that of recalling a certain story given that it was reported as a reminding ($p(\text{recall}|\text{reminding}) = .87$) and (b) that of reporting a certain story as a reminding given that it was reported in free recall ($p(\text{reminding}|\text{recall}) = .79$). The number of study stories each subject had

available in memory was calculated as the total number of different stories reported as recalls or reminders (or both); this mean was 5.5.

Table 3 shows the mean proportion of matches and mismatches for each story type. A *t* test of matches for complete-theme versus predict-theme stories was significant with subjects as a random factor, $t(53) = 3.8, p < .0001$, and showed a trend toward significance with materials as a random factor, $t(14) = 1.76, p < .10$. *t* tests of mismatches for the two story types showed a trend toward significance under a subjects analysis, $t(53) = 1.64, p < .11$, but was not significant in a materials analysis.

To determine whether the number of responses to complete-themes and to predict-themes differed from chance, these means were tested against two estimates of chance performance. First, subjects' reported complete-theme and predict-theme matches were corrected by a factor based on the probability of reporting any story from the study set (.125) and the number of responses each subject actually made in the test phase of the experiment (whole set method). Both complete-theme matches and predict-theme matches were significantly different from this chance estimate, $t(54) = 10.07$ and $t(54) = 8.22$, respectively; $p < .0001$ for both. For complete-theme stories, subjects' scores on average exceeded the score expected by chance by 1.27 observations; for predict-themes, the mean was .88 observations. A second estimate, looking at performance only on trials where the subject had the matching story available in memory, with "available" de-

finied as a story reported as a recall or reminding (memory set method), also showed that both complete-themes and predict-themes were matched at higher-than-chance levels, $t(53) = 8.17$ and $t(53) = 7.57$, respectively; $p < .0001$ for both.

To test the rate of mismatches and nulls across all three story types (complete, predict, filler), two 1×3 analyses of variance were performed. The story types did not differ in the proportion of mismatch responses when analyzed either by subjects or materials, but did differ in proportion of null responses under both analyses, $\min F'(1,23) = 16.55, p < .0005$, with filler stories showing the highest proportion of null responses. The mean number of nulls for fillers, predict-themes, and complete-themes were .89, .54, and .35, respectively.

Discussion

The results show that one can reliably access previous cases in memory using a partial set of abstract features. Both of the chance measures used show that partial stories lead to appropriate reminders at a level higher than expected by chance. Because the estimates of chance performance assumed a response for each trial, they represent a conservative measure, as subjects often left spaces blank. These estimates may therefore underestimate the degree of access provided by a partial-features condition. The complete-theme condition replicates the finding that abstract similarities in thematic patterns of features can be reliably used to retrieve related cases from memory (Gentner & Landers, 1985; Rattermann & Gentner, 1987; Seifert et al., 1986).

The low response rate for filler stories shows that subjects could discriminate thematically related stories from unrelated ones and that both the complete-theme stories and the predict-theme stories contained enough information to allow subjects to do so. Further, the low mismatch rates for both types of thematic stories show that the matches were not simply due to guess-

TABLE 3
AVERAGE PROPORTION OF RESPONSES BY ALL STORY VERSIONS, EXPERIMENT 1

Response type	Story type		
	Complete-theme	Predict-theme	Filler
Match	.48	.33	—
Mismatch	.17	.13	.11

ing; if so, many more spurious reminders might be expected than were observed.

Conditional probability results provide some evidence that subjects were retrieving cases from memory based on the test cues and not just using a strategy of keeping the whole study set in mind and serially comparing each study story to the test story. With a strategic search, one would expect subjects to recall all the items they put down as reminders and vice versa, which would result in conditional probabilities of 1.0. However, the probability of a reminding given free recall of the story later was .79, suggesting that at least some reminders were not drawn from a search set. The final free recall is at best ambiguous evidence of an initial search set, as subjects did this task last, under no time constraints, and the stories retrieved from memory during the reminding task may have remained accessible and thus inflated the recall score. Further, subjects were only exposed to the study stories for 5 min, which allowed most to read through each story only once, and limited further rehearsal, as did the subsequent distractor task.

The finding that complete stories resulted in more matches than the partial stories did is as expected: The complete stories provide more retrieval cues, which would be more likely to lead to access, and would also allow more complete reconstruction of the original story. As such, the complete-theme condition seems to place an upper bound on the potential effectiveness of partial feature sets.

In summary, these results indicate that people can use abstract features to access previous material, even when given only a portion of the complete feature set. Such features could lead to reminders when the most important similarities between two cases are abstract ones, and in general, use of partial feature sets is particularly important in a planning or problem-solving context, where one would not have a complete set of features to work with. Given these results, we can now look at whether, in

cross-contextual situations where information is incomplete, different types of abstract feature sets result in better access to related prior cases.

EXPERIMENT 2

If functionality is an organizing memory principle for structures like TAUs, as case-based reasoning theory suggests (Schank, 1982), then features in an input that predict a failure should form a privileged set that leads to more reliable access than would be attained using other sets of abstract features. Optimally, one should access a relevant theme after receiving information critical to the planning decision, but before deciding upon and taking an action.

To determine whether predictive features provide an advantage in retrieval, the reminding paradigm was used. As in Experiment 1, subjects received a set of study stories first, but in this experiment all the test stories were incomplete exemplars. Half of the stories (predict-themes) presented only elements leading up to a decision point, whereas the other half (theme-outcomes) only contained a planning decision and an outcome. The predictive-features hypothesis predicts that the most useful time to retrieve the study story is before the decision is made and executed. Thus, stories containing features that predict bad outcomes for certain planning decisions should bring to mind past instances of the failure. Reminders based on other abstract feature sets, however, would not provide one with the opportunity to avoid a planning failure and so would be less functional. Thus, under a predictive features hypothesis, one would not expect privileged access based on theme-outcome stories, which provide only the outcome of an incident, too late to benefit current planning.

Method

Subjects. Fifty-four University of Michigan undergraduates participated in a single session lasting about 1 h. They received

course credit in an introductory psychology class for participating and were run in groups of six to ten.

Materials. The materials consisted of a study packet, a test packet, and a distractor task. The study packet and distractor task were the same as those used in Experiment 1. Ten study packets, each containing the same eight stories in random order, were used.

The test packets, as in Experiment 1, contained eight test stories, each of which shared a theme with one of the study stories, and four fillers, which were based on adages not represented in the study stories. All 12 test stories had two versions: a predict-theme version (60–70 words), which included the initial elements of a thematic pattern but not the conclusion or resolution of the story, and a theme-outcome version (40–50 words), which presented only the planning step taken and outcome of the story. See Table 1 for sample predict-theme and theme-outcome test story versions.

The abstract features characterizing an example target theme, and the elements its test and study story contained, are presented in Table 2. The predictive features chosen were those that one could perceive before the decision point in a planning situation and that would reasonably lead one to expectations about what might occur next. The theme-outcome features were the remaining elements, that is, the planning decision and its consequences. The test stories were partitioned into mutually exclusive feature sets, to eliminate elements common to both versions.

Each subject received four target predict-theme stories, four target outcome-themes, and four fillers (two in predict-theme format and two in outcome-theme format). The counterbalancing and randomization of stories, and the instructions after each of the test stories, were the same as those in Experiment 1.

Procedure. The procedure was the same as that in Experiment 1, except that the instructions for the test packet and the free

recall task were written, as opposed to being delivered verbally.

Results

Responses were coded using the same criteria as those in Experiment 1, for both subjects and materials analyses. Only 2.59% of the responses occurred as the second of a double reminding. Responses that could not be identified as one of the target stories (intrusions) accounted for 5.5% of the responses.

Subjects wrote down a mean of 3.67 responses as reminders and could recall, on average, 4.74 study stories in the free-recall phase. As in Experiment 1, the overlap of these two measures was high. The average probability of reporting a certain story as a free recall given that it was reported as a reminding ($p(\text{recall}|\text{reminding})$) was .89, and the mean probability of reporting a certain story as a reminding given that it was recalled at the end ($p(\text{reminding}|\text{recall})$) was .71. The mean number of study stories subjects had available in memory (the number of different stories that were reported as recalls or reminders or both) was 5.06.

A *t* test of the number of matches in response to predict-theme versus theme-outcome stories was not significant with either subjects or materials as a random factor. However, mismatches in response to theme-outcome stories occurred significantly more often than they did in response to predict-themes, $t(53) = 2.99$, $p < .004$; $t(14) = 2.44$, $p < .029$, for subjects and materials analyses, respectively. Table 4 shows the mean proportion of matches and mismatches for each story type.

The mean number of responses to predict-theme stories and to theme-outcomes was tested against the two chance estimates used in Experiment 1. Using the whole set method, where subjects' reported predict-theme matches and theme-outcome matches were corrected by a factor estimating chance performance, predict-theme matches occurred at a rate significantly higher than the chance estimate, $t(53) =$

TABLE 4
AVERAGE PROPORTION OF RESPONSES BY ALL STORY
VERSIONS, EXPERIMENT 2

Response type	Story type		
	Theme-outcome	Predict-theme	Filler
Match	.30	.35	—
Mismatch	.20	.10	.10

7.33, $p < .01$. Theme-outcome matches, however, showed only a trend towards a difference, $t(53) = 2.02$, $p < .10$. For the predict-theme condition, the average number of observations exceeded the chance estimate by .500; for outcome-themes, this figure was .204. The memory set method, which looked at performance only on trials where the subject had the matching story available in memory, showed that both predict-themes and theme-outcomes were matched at a higher-than-chance level, $t(53) = 6.24$ and $t(53) = 5.46$, respectively; $p < .001$ for both.

Discussion

The results of this experiment suggest that predictive features show an advantage in terms of selectivity of reminding. The predict-theme and theme-outcome stories did not differ in number of matches, which suggests that they both provide reliable access to related cases; however, they did differ in number of mismatches generated. This could indicate that the elements in the predict-theme stories distinguished the themes more clearly, and thus subjects tended either to find the right story or to give no answer. The theme-outcomes, however, tended to evoke a wider range of cases as responses, indicating that the features they provide were shared by other potentially retrievable episodes. Such outcome cues may not have the specificity found in the initial predictive elements and thus could lead to spurious matches as well as accurate ones. For example, several TAUs deal with resource allocation; one

might access any of them given only information that "resources were wasted."

The comparison of responses to predict-theme stories and theme-outcomes using the whole set chance estimates supports this view. This chance estimate looks at how likely subjects are to discriminate the thematic stories from the fillers and, having done that, to pick the "right" answer from the chosen stories. These estimates show that subjects matched predict-theme stories at a higher-than-chance rate, whereas they responded to theme-outcome stories at closer to chance levels. Subjects put down a greater number of inappropriate responses to theme-outcomes.

However, when using the memory set chance measure, subjects reported both predict-themes and theme-outcomes at rates significantly better than chance. This measure looks at only those test stories that the subject, given the stories he or she had in memory, could potentially have matched correctly. For example, if the study story about Mike were the appropriate response, but the subject did not report that story in either the reminding or the free-recall portion of the experiment, the corresponding test story would not be included among the set that the subject could have potentially matched. Thus, any response given for that test story would be ignored, rather than being counted as an error (as would occur in the whole set method). Comparing these two estimates, it appears that if one had the appropriate theme-outcome available in memory, one could put it where it belonged, as with the predict-themes. However, if one did not have the right response available, theme-outcomes allowed access to other exemplars that might do, whereas predict-themes, being more specific, led to either matching or no response.

Thus, Experiment 2 provides evidence of access to prior cases based on both types of feature sets. However, the evidence shows predict-theme features lead to more reliable access, perhaps because of the specificity in matching they provide compared to

theme-outcome features. This distinctiveness property is examined further in Experiment 5. However, an alternate explanation for these results may lie in a differential degree of similarity between the cue types and the target stories. Experiments 3 and 4 were designed to determine whether similarity alone could account for these differences in accessibility.

EXPERIMENT 3

The advantage for predict-themes in access to prior cases has been explained in terms of the functional purpose of the reminders. An alternative explanation for this finding is that the two feature sets differ in overall similarity to the study stories and that this alone accounts for differences in access. To determine whether the thematic test stories in the two conditions were differentially similar to the study stories, subjects rated the overall similarity of the pairs of stories used in the previous two experiments. The comparisons included thematically related stories and filler stories, either in complete-theme, predict-theme, or theme-outcome versions. If the results in the ratings task, where subjects need not rely on memory for the stories, show that subjects judge predict-themes as more similar overall to study stories than theme-outcomes are, this would suggest that the differences can be explained by a greater similarity to the targets in the predict-theme condition, rather than by any predictive or functional role they may play. However, if predict-themes and theme-outcomes do not differ in overall similarity, it would support the claim that the differences found between predict-theme and theme-outcome test stories were due to their differing effectiveness as distinctive memory retrieval cues.

Method

Subjects. The subjects were 60 University of Michigan undergraduates who did this task as part of another experiment.

Subjects were paid for their participation and were run in groups of 18 to 20.

Materials and procedure. The materials consisted of the eight study stories used in Experiments 1 and 2, their corresponding thematic test stories, and the four filler test stories used in both experiments. Each thematic test story and filler story had three versions: the complete-theme version used in Experiment 1, the predict-theme version used in Experiment 2, and the theme-outcome version used in Experiment 2. Each study story was paired with its match (using all three versions) and with a filler chosen at random (also using all three versions), for a total of six rating conditions for each story. Each of the four fillers was paired with two different study stories at random.

Each study story occurred only once in each ratings packet, being paired with one of its six test stories. The test story pairings were counterbalanced in a Latin square design, with two constraints: (1) that a particular version type (complete-theme, predict-theme, or theme-outcome) occurred no more than three times in any packet, and (2) that each packet contained four filler and four thematic stories. This last constraint prevented any filler from appearing twice in the same ratings packet. Within each packet the pairs of stories were placed in random order. Each ratings page in the packet featured a complete study story, followed by its paired test story. Following this was a scale, ranging from 1 to 7, on which the subject rated the overall similarity of the stories, with 1 being "not very similar" and 7 being "very similar." Each page contained only one story pair and rating scale. The instructions were to rate the overall similarity of the pairs of stories, working forward through the packet.

Results

The ratings were analyzed in a 2×3 analysis of variance with type of story (thematic or filler) and story version (complete-theme, predict-theme, or theme-outcome)

as factors. The average ratings for each condition are shown in Table 5.

The results showed a significant main effect of story type for both subjects and materials, $\min F'(1,37) = 93.16, p < .0001$, with thematic stories receiving higher ratings. Story version also showed a main effect, with complete-theme stories receiving the highest ratings overall, followed by theme-outcomes, and then by predict-themes. This was significant under both subjects and materials analyses, $\min F'(2,16) = 6.26, p < .01$.

Further planned comparisons tested the difference between the predict-theme and theme-outcome conditions for thematic stories and filler stories separately. The analysis of thematic stories showed a trend under both subjects and materials analyses, $t(58) = 1.88, p < .065$ and $t(7) = 1.99, p < .087$, respectively, with theme-outcome stories being judged more similar to the study stories than predict-theme stories were. For filler stories, this result was significant under both subjects and materials analyses, $t(58) = 3.22, p < .002$; $t(7) = 3.68, p < .008$, respectively.

Discussion

Most importantly, the results suggest that the memory advantage of predict-themes cannot be explained by a greater overall similarity to the study stories compared to the theme-outcomes. In fact, subjects rated theme-outcomes as having *more* overall similarity to the study stories using this direct comparison rating method in-

stead of a reminding task. That theme-outcome stories were rated higher than predict-theme stories is mainly accounted for by differences among the filler story ratings (see Table 5). With the filler stories, subjects' ratings may reflect vague similarities to the target case, such as "sharing negative outcomes." Noticing vague similarities could result in higher ratings in direct comparison, but could lead to unreliable and spurious remindings in a task involving memory access. For story version differences, the finding that complete-theme stories had the highest ratings is not surprising because they contained more information that could potentially enter into a match.

EXPERIMENT 4

Experiment 3 looked at the alternative explanation that overall similarity accounts for the advantage for predict-themes in access to prior cases. However, one could still argue that the predict-theme stories allow more reliable access because they have more specifically *thematic* features in common with the study stories than theme-outcome stories do. To determine whether the thematic test stories in the two conditions were differentially similar to the study stories at this level, subjects rated the thematic, rather than the overall, similarity of the pairs of stories used in Experiments 1 and 2, using the same procedure as in Experiment 3. If subjects can reliably detect thematic similarities in all the thematic conditions, it would suggest that the differences found between predict-theme and theme-outcome test stories were due to their differing effectiveness as memory retrieval cues. On the other hand, if the results in the ratings task, where subjects need not rely on memory for the stories, show that subjects judge predict-themes as more thematically similar to study stories than theme-outcomes are, this would suggest that the differences can be explained by their having greater similarity to the targets.

TABLE 5
AVERAGE RATINGS OF STORIES BY STORY TYPE AND
VERSION, EXPERIMENT 3

Story version	Test story type	
	Theme	Filler
Complete	5.25	2.77
Theme-outcome	4.61	2.78
Predict-theme	4.03	1.93

Method

Subjects. The subjects were 21 University of Michigan undergraduates who did the ratings as a subtask embedded within another experiment. They received course credit in an introductory psychology class for their participation. Subjects were run individually.

Materials. The materials used were the same as those in Experiment 3, except for the instructions and the rating scales. These were based on those used by Gentner and Landers (1985), with a few modifications to emphasize use of thematic features regardless of story length. Subjects were told that they needn't see the whole story to get the point and to see its similarity to other stories. This was done to focus subjects on the stories' content and limit their use of the stories' length as a factor in their ratings. Each page in the ratings packet featured a complete study story, followed by its paired test story. Following this was a scale, ranging from 1 to 6, on which the subject rated the soundness of the match between the stories, with 1 being "unrelated" and 6 being "extremely sound." This scale was used, rather than one stressing similarity, to focus subjects on just the thematic similarities of the stories and away from overall similarity, as assessed in Experiment 3. Each page contained only one story pair and rating scale.

Procedure. Subjects were told to read the instructions and fill out the pages in the ratings booklet.

Results

The ratings were analyzed in a 2×3 analysis of variance with type of story (thematic or filler) and story version (complete-theme, predict-theme, or theme-outcome) as factors. The average ratings for each condition are shown in Table 6.

The results showed a significant main effect of story type for both subjects and materials, $\min F'(1,24) = 41.59, p < .0001$,

TABLE 6
AVERAGE RATINGS OF STORIES BY STORY TYPE AND VERSION, EXPERIMENT 4

Story version	Test story type	
	Theme	Filler
Complete	4.98	2.67
Theme-outcome	4.23	2.16
Predict-theme	4.04	1.51

with thematic stories receiving higher ratings. Story version also showed a main effect, with complete-theme stories receiving the highest ratings overall, followed by theme-outcomes, and then by predict-themes. This was significant under both subjects and materials analyses, $\min F'(2,35) = 4.945, p < .025$.

Further planned comparisons tested the difference between the predict-theme and theme-outcome conditions for thematic stories and filler stories separately. The analysis of thematic stories showed no significant effect of version (predict-theme versus outcome-theme) with either subjects or materials as random factors. For filler stories, there was a trend toward theme-outcome stories being judged as more similar to the test stories than predict-theme stories were, but only under subjects analysis, $t(29) = 1.78, p < .085$.

Discussion

The results of this experiment mirror those of Experiment 3, and most critically, show that the memory advantage of predict-themes cannot be explained due to predict-themes having greater thematic similarity to study stories than theme-outcomes do. The main effect of story type supports the claim that each story version, complete-theme, predict-theme, and theme-outcome, contained cues that were sufficient to identify the main point, or theme, of the story as a whole, and allowed subjects to detect a high degree of similarity with the study story. The lack of difference between thematic predict-themes and theme-outcomes

also suggests that the features they contain characterize the themes equally well under direct comparison. Thus, predict-themes were not rated as more similar to the target study stories than theme-outcomes were, ruling out a thematic similarity explanation for the predict-theme access advantage in Experiment 2.

EXPERIMENT 5

Experiments 3 and 4 found no difference in the degree of similarity of predict-themes and outcome-themes to the study stories they matched. However, for distinctiveness to explain the access advantage for predict-themes, each set of predictive features would have to be highly similar to its paired study story and relatively dissimilar to any others in the set, and outcome themes would have to be similar to several stories in the set. To further test this hypothesis, Experiment 5 used a direct comparison method, in which subjects were allowed to look at the study stories rather than recall them from memory and were asked to write down study stories that matched the test stories. If the same pattern of results occurs as that in Experiment 2, this would further support a distinctiveness explanation for the utility of predictive features.

Method

Subjects. Forty-seven University of Michigan undergraduates participated in a single session lasting about 1 h. They received course credit in an introductory psychology class for participating. Subjects were run in groups of 20 to 30.

Materials. The materials consisted of the study packet and test packet used in Experiment 2, with slightly modified instructions. Subjects were told to "match" the stories in the second packet to those in the first, and that they could look through the study packet as often as they liked, but were to go forward through the test packet and not re-

turn to pages already completed. Subjects were not told what criterion to use in determining whether any study stories "went with" each of the test stories. Subjects were not asked to make "thematic matches" (the evidence for subjects' ability to match thematic stories was established in Seifert et al., 1986); instead, this study measures whether the matches that subjects make *without any memory demand* are in fact thematic matches.

Procedure. The procedure was the same as that in Experiment 2, except that there was no distractor task between when subjects read the study stories and when they received the test packet.

Results

Responses were coded using the same criteria as those in Experiments 1 and 2, for both subjects and materials analyses. Only one response could not be identified as one of the target stories (intrusion).

Subjects wrote down a mean of 7.2 responses in the test packet. A within *t* test of number of double responses by story type (predict or outcome) showed a significant difference, $t(46) = 3.03$, $p < .004$ for subjects and $t(7) = 2.77$, $p < .028$ for materials, with means of .17 for predict-themes and .53 for outcome-themes.

A *t* test of the number of matches in response to predict-theme versus theme-outcome stories was not significant with either subjects or materials as a random factor. However, mismatches in response to theme-outcome stories occurred significantly more often than they did in response to predict-themes, $t(46) = 3.72$, $p < .001$ with subjects as a random factor. A materials analysis showed a trend in the same direction, $t(14) = 1.72$, $p < .11$. These results are shown in Table 7.

The mean number of responses to predict-theme stories and to theme-outcomes was also tested against the whole set chance estimate used in Experiment 1, where subjects' reported predict-theme

TABLE 7
AVERAGE PROPORTION OF RESPONSES BY ALL STORY
VERSIONS, EXPERIMENT 5

Response type	Test story type		
	Theme-outcome	Predict-theme	Filler
Match	.48	.52	—
Mismatch	.28	.13	.40

matches and theme-outcome matches were corrected by a factor estimating chance performance. Both predict-theme and outcome-theme matches occurred at a rate significantly higher than the chance estimate, $t(47) = 14.26$ and $t(47) = 24.49$, respectively, $p < .0001$ for both. The two conditions averaged 1.3 and 1.51 observations in excess of chance, respectively.

Discussion

The results of this experiment support distinctiveness as an explanation of the findings in Experiment 2; specifically, that predict-theme cues are more distinctive, yet not more similar, in regard to the study stories than are outcome-theme cues. In Experiment 5, outcome-theme stories were more than twice as likely to result in multiple responses per story, and the results again showed that, although both types of stories resulted in a large number of matches, outcome-theme stories received more mismatch responses than predict-theme stories did. Thus, whereas the predictive and the outcome elements of a story are equally similar to the target case, the two feature sets differ in how uniquely they characterize elements that discriminate the target from distractors.

The finding that matches occurred at a rate much higher than chance for both types of feature sets, and the overall high number of responses, is consistent with the fact that the task involved direct comparison of similarities and not comparison from memory. This direct method may encourage responding based on vague similarities

that would not be detected in a memory task, as evidenced by the relatively high rate of responses to filler stories.

GENERAL DISCUSSION

The results of the experiments presented here provide support for a functional approach to the organization of cases in memory, as suggested by Schank (1982). A functional analysis suggests that for one to benefit from potentially useful abstract similarities between cases when planning, one *must* have access to prior cases and the planning information they contain before a complete thematic structure (defined as themes from cultural adages) is evident in the new situation; otherwise one will not benefit from memory of prior mistakes before the outcome is determined. Thus, features in the input that predict potential problems, as opposed to other sets of abstract features, should lead to more reliable access to failure information, as proposed by the predictive-features hypothesis.

In the first two experiments, stories with incomplete thematic structures led to reminders at a rate much higher than that expected by chance, suggesting that one need not receive an entire thematic pattern before accessing a related case in memory. As might be expected, having a complete pattern results in better retrieval because this potentially allows the most complete match with material stored in memory. However, the fact that the number of matches based on incomplete feature sets significantly exceeded the number of spurious reminders suggests that, for the most part, reminders potentially useful to a planner occurred.

The results also support the hypothesized advantage of predictive features over outcome features in recall. In planning situations, the most useful similarities would be those that let one anticipate potential problems before they occur. In a functionally organized memory, such predictive cues could be used to index knowledge

structures that contain planning information on how to avoid potential pitfalls. The results are consistent with case-based reasoning models (Hammond, 1989; Kolodner, 1983; Riesbeck & Schank, 1989; Schank, 1982) which propose that one tends to encode the goal-relevant or predictive features of an episode, as inferred during initial comprehension. These features are then used to index the episode within memory. Then, if one discovers the same features while understanding a new case, previous instances sharing those indices will be retrieved as a result of this functional organization of memory.

The experimental results presented here suggest that, under such a memory organization, clusters of partial abstract features that predict previously experienced planning failures lead not to *more*, but to more *reliable* access to relevant previous instances than do other abstract, equally similar associated features. When provided with just outcome-theme features, subjects accessed more mismatches than they did with predictive features, suggesting that the outcome features do not provide information that is as specifically tied to a particular theme. Outcomes of planning decisions may share broader abstract commonalities such as "plan successful" and "plan fails" and therefore would not identify any one specific planning scenario over any other. Thus, the perceived features that are likely to index into prior related cases will be exactly the features that detect important factors in making the planning decision and that distinguish this case from other, related planning cases. That is, people will learn the features that are helpful in deciding what to do (and what not to do) and that separate that type of decision from other related types. More generally, the claim is that, although one can encode cases and retrieve them from memory based on a variety of similar features, one will most often and most successfully utilize the set that includes the causal features that identify the specific planning problem, both in

terms of grouping similar cases together and of distinguishing them from other cases with other similarities.

An alternative explanation for these results is that the predictive and outcome stories differed in their thematic similarity to the original study stories. However, the results of Experiments 3 and 4 rule out this interpretation, because subjects rated the two types of incomplete thematic stories as equally similar to the source stories. This suggests that the advantage for the predictive features is not that they are more similar overall, nor that they contain more thematic information and are therefore more similar to prior cases, but that the predictive features provide more *distinctive* access to directly relevant cases. The results of Experiment 5 further support a distinctiveness explanation, as more spurious reminders occurred in response to theme-outcome stories than predictive ones in direct (rather than memory) comparison as well. The results suggest that different kinds of abstract similarities can affect access based on factors other than their overall or thematic similarity to a previous, "best-matching" case in memory. Distinctiveness, rather than similarity, is the critical determinant of retrieval when the pool of potential matches increases to several candidates that may match equally well. Distinctiveness is certainly not a new principle in theories of memory retrieval (see Crowder, 1976), but its importance has been overlooked in previous work on analogy, perhaps because of the tendency to study analogical relationships that are far removed from other cases in memory (Hammond, Seifert, & Gray, 1991).

A potentially confounding factor is that the predictive features are always placed in the first part of the stories used, suggesting that some sort of primacy effect, and not the nature of the features themselves, could account for our results. If subjects have better memory for the initial elements of the study stories, then they should find matching the predict-theme test stories eas-

ier than matching the theme-outcomes. However, it is not clear that primacy effects would occur within individual stories, as opposed to within the set of stories. Subjects may have better recall for whole stories that they encountered first in the study packets, but this in itself would not favor predictive features over outcomes. Further, if the unit of analysis is the individual story, one might expect both primacy and recency effects, with the middle of the story forgotten. Again, this would not favor predictive features over outcomes.

The reminding methodology used in the first two experiments may actually underestimate the number of cases one can access with incomplete patterns. The experimental task required subjects to write down a story that *matched* the test story. Subjects may have been reminded of a case but felt it was inappropriate to report as a match per se; this was more likely to occur in the incomplete story conditions, where less information was available to identify a theme with certainty. Editing reminders seems common, especially in tasks like conversation, where appropriateness is an important consideration. In a problem-solving task, defined as such, subjects may more willingly report and use reminders based on incomplete thematic patterns.

These findings also provide further evidence for the validity of thematic knowledge structures in human memory (Dyer, 1983; Seifert et al., 1986) and they extend work on the role of structural features in analogical access (Gick & Holyoak, 1980; Holyoak, 1985; Ross, 1989a) by showing that even incomplete patterns containing such features can be effective. The focus on complete feature sets in prior research on access to analogies may result in failing to detect the patterns of similarities that people can utilize in planning and problem solving. For example, in materials drawn from Gentner and Landers (1985), two stories may share many first-order relations early on (Gentner, 1989) yet have different endings. For example, a bird gives feathers

to a hunter and later is either shot with an arrow fletched with its own feathers, or spared when recognized as the hunter's benefactor. Taken in the context of whole stories, these two outcomes are not analogous; however, if considered only up to the decision point (whether to give the feathers), the two stories could be considered analogous, and in a planning situation both could lead to valuable reminders (Hammond et al., 1991). The feature sets characterizing abstract themes are not simply unordered clusters of shared features; instead, they are organized by causal links. Thus, predictive features stand in a particular causal relationship to the conclusory outcomes.

The findings also broaden the view of analogy in suggesting that the set of all commonalities one can detect in post hoc comparisons is not necessarily the same as the set of commonalities available for use in memory retrieval. For example, the applicability of the convergence solution to both the General and the Ray problem (Gick & Holyoak, 1980) certainly contributes to the high overall similarity when in a rating task, but since that solution is not present in the Ray problem as presented, it can be of little use in access. Although Gick and Holyoak (1980, 1983) found a low rate of reminders without a retrieval hint (about a third of the subjects), the ones that did occur would have involved matching based on partial features. For problem-solving settings, a less-than-complete set of matching features will be available, requiring a process model of analogy that can account for access involving incomplete comparisons.

One could argue that the distinctiveness effect occurs in the experimental materials but might not generalize beyond them. However, to eliminate the effect found here, one would have to find themes with relatively confusable predictive elements and much more specific outcomes. This could be difficult, because for a set of features to actually be predictive, they need to pick out a certain outcome and not be as-

sociated with many outcomes. Confusability of predictive features would diminish their usefulness as predictors (by definition). Although there may be some distinctions among abstract features that characterize outcomes, there are likely to be fewer different kinds than there are potential distinctions among predictive features, which include conditions and constraints involved in selecting an appropriate plan. Any feature set with similarly defined predictive power will lead to more reliable (though not necessarily more) access. What is, in fact, predictive could vary by domain and by task, as far as what one needs to predict, when, and what information is available at the time. Thus, predictive power is only one of many possible forms of distinctiveness within a set of cases. It is simply the specific relationship of information contained in predictive features to the prediction that needs to be made that is the reason for their advantage in selecting the most helpful prior cases in memory.

In summary, the results provide evidence that one can access cases in memory that share abstract features, even when given incomplete cues. Thus, one can access valuable, cross-contextual planning information *before* observing a total set of matching abstract features. From a functional perspective, the most useful of these incomplete feature sets will contain elements that *predict* potential problems, and allow access to relevant cases before one must make a planning decision, thereby improving the planner's ability to predict and avoid planning failures. These findings are consistent with work in case-based reasoning (Hammond, 1989; Kolodner, 1983; Riesbeck & Schank, 1989; Schank, 1982), which proposes that planners use modifications of analogous past experiences to solve problems in addition to problem-solving rules. These results also suggest a functional approach to indexing for use in planning systems. Further research is needed to discover functional feature sets appropriate to other tasks and domains, to extend the

general principle of functionality in memory indexing.

APPENDIX

Key: A letter by a story element means that the story version contains that element.

- S Study story
- C Complete-theme test story
- P Predict-theme test story
- O Theme-outcome test story

Theme: Blind Leading the Blind

Elements:

- S C P O (1) X has goal.
- S C P (2) X needs other resources to achieve it.
- S C P (3) Access to source A readily available.
- S C P (4) X fails to take into account A's reliability.
- S C O (5) X accepts A's advice.
- S C O (6) X executes recommended plan.
- S C O (7) X fails to achieve goal.

Theme: Don't Throw Good Money after Bad

Elements:

- S C P O (1) X has goal.
- S C P (2) X invests resources in plan.
- S C P (3) Plan fails.
- S C P (4) X does not assess reasons for failure.
- S C O (5) X invests more resources in same plan.
- S C O (6) Plan again fails.
- S C O (7) X is left without resources.

Theme: Counting Your Chickens before They've Hatched

Elements:

- S C P (1) X desires A.
- S C P (2) X assumes A.
- S C P (3) X does not act to ensure or verify A.
- S C O (4) X invests resources based on A.
- S C O (5) A does not occur.
- C O (6) X has used up resources for nothing.

Theme: Too Many Cooks Spoil the Broth

Elements:

- S C P O (1) X and Y have the same (token) goal.
- S C P (2) X and Y have plans.
- S (3) X and Y's plans together will miss goal.

- S C P O (4) X and Y do not communicate to coordinate actions.
 (5) X and Y execute plans independently.
 S C O (6) Interaction of plans does not achieve goal.

Theme: Cutting off Your Nose to Spite Your Face

Elements:

- S C P O (1) X has goal.
 S C P (2) X needs additional resources from source A.
 S C P (3) Source A does not provide resources.
 S C P O (4) X uses force on A.
 S C O (5) Force has negligible effect on A.
 S C O (6) Force destroys resources X has already.

Theme: The Cure is Worse than the Disease

Elements:

- S C P O (1) X has goal.
 S C P (2) X acts to achieve goal.
 S C P (3) Plan falls short of goal but difference between present state and goal is small.
 S C O (4) X acts to minimize difference.
 S C O (5) Action results in a state worse than the previous state.

Theme: Closing the Barn Door after the Horse is Gone

Elements:

- S C P (1) X has goal.
 S (2) X is aware of potential goal blockage.
 S C P (3) X has knowledge of solution.
 S C P O (4) X has power to execute solution.
 S C P (5) X does not take action.
 S C P O (6) Goal gets blocked.
 S C O (7) Solution conditions change.
 S C O (8) X executes original solution anyway.
 S C O (9) Executed solution doesn't work.

Theme: Pot Calling the Kettle "Black"

Elements:

- S C P (1) X states an extreme opinion.
 C P (2) X has motive for contradiction.
 S C (3) X has opportunity for contradiction.
 S C (4) X acts contrary to stated opinion.

REFERENCES

- ANDERSON, J. R. (1986). Knowledge compilation: The general learning mechanism. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 289-310). Los Altos, CA: Kaufman.
- COLLINS, G. C. (1987). *Plan creation: Using strategies as blueprints*. Unpublished doctoral dissertation, Yale University.
- CROWDER, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- DYER, M. G. (1983). *In-depth understanding: A computer model of integrated processing for narrative comprehension*. Cambridge, MA: MIT Press.
- GENTNER, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- GENTNER, D. (1989). Finding the needle: Accessing and reasoning from prior cases. *Proceedings of the Second DARPA Case-based Reasoning Workshop*. Pensacola, FL.
- GENTNER, D., & LANDERS, R. (1985, November). Analogical reminding: A good match is hard to find. *Proceedings of the International Conference on Systems, Man, and Cybernetics*. Tucson, AZ.
- GICK, M. L., & HOLYOAK, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
- GICK, M. L., & HOLYOAK, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- HAMMOND, K. (1989). *Case-based planning: Viewing planning as a memory task*. San Diego: Academic Press.
- HAMMOND, K., SEIFERT, C. M., & GRAY, K. (1991). Functionality in analogical transfer: A hard match is good to find. *Journal of the Learning Sciences*, 1(2), 111-152.
- HOLYOAK, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59-87). New York: Academic Press.
- KOLODNER, J. L. (1983). Reconstructive memory: A computer model. *Cognitive Science*, 7, 281-328.
- LEHNERT, W. (1980). Plot units and narrative summarization. *Cognitive Science*, 5, 293-331.
- LEHNERT, W., BLACK, J. B., & REISER, B. J. (1981). Summarizing narratives. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B.C.
- NOVICK, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 510-520.
- PIROLI, P. L., & ANDERSON, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology*, 39, 240-272.

- RATTERMANN, J. J., & GENTNER, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. In J. Anderson (Ed.), *Proceedings of the Ninth Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- REISER, B. J., & FARIES, J. M. (1988). *Access and use of previous solutions in a problem solving situation* (Tech. Rep. No. 29). Princeton: Princeton University Cognitive Science Laboratory.
- RIESBECK, C. K., & SCHANK, R. C. (1989). *Inside case-based reasoning*. Hillsdale, NJ: Erlbaum.
- ROSS, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629-639.
- ROSS, B. H. (1989a). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 456-468.
- ROSS, B. H. (1989b). Reminders in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.
- SCHANK, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people*. New York: Cambridge Univ. Press.
- SEIFERT, C. M., DYER, M. G., & BLACK, J. B. (1986). Thematic knowledge in story understanding. *TEXT*, 6(4), 393-426.
- SEIFERT, C. M., MCKOON, G., ABELSON, R. P., & RATCLIFF, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 220-231.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- VAN DIJK, T. A. (1979). *Macrostructures*. Norwood, NJ: Ablex.

(Received August 21, 1991)

(Revision received December 10, 1991)