PREDICTIVE UTILITY IN CASE-BASED MEMORY RETRIEVAL

Hollyn M. Johnson and Colleen M. Seifert
University of Michigan

Abstract

The problem of access to prior cases in memory is a central issue in the case-based reasoning paradigm. Previous work on thematic knowledge structures has shown that using a complete exemplar of a thematic pattern allows access to the structure and related cases in memory. However, the knowledge and expectations provided by such structures can aid in planning and problem-solving. Therefore, to be most useful, the information should become available *before* the input pattern is complete. Retrieval must therefore be possible based on only a subset of the features present in the full thematic pattern. This study investigated whether a pattern that contains elements predicting an outcome, but not the outcome itself, would result in access comparable to that found when a full pattern is used. The results showed that subjects were less successful accessing the thematic structure using partial patterns than they were when using full patterns. However, remindings based on partial patterns occurred more often than would be expected by chance. We conclude that partial patterns contain some predictive features that can allow access to a thematic knowledge structure before the pattern is complete.

Introduction

A central problem for models of case-based reasoning is the retrieval of appropriate cases from memory. Retrieval is assumed to depend on matching an input to the contents of memory, which is easier if the new case is highly similar to a case already stored, and highly distinct from other cases in memory. Two cases may be similar on different levels of abstraction; they may share surface features, such as characters and settings, or contexts, or, at the highest level of abstraction, themes. In planning and problem solving, it would be useful to access cases in memory that share abstract features with the input case because those cases could provide valuable, cross-contextual planning information. In a planning situation, optimal retrieval of past cases would be based on features that predict potential planning failures and would occur *before* the outcome of the situation has been determined. The issue, therefore, is how much information is needed to access past cases.

Recent research on reminding has focused on the role of highly abstract, thematic knowledge structures in access to individual cases (Seifert, McKoon, Abelson, and Ratcliff, 1986). The content of such abstract structures has been characterized by Lehnert (1980) as "plot units", Schank (1982) as thematic organization points (TOPs), and by Dyer (1983) as thematic abstraction units (TAUs). Such structures capture more abstract relationships between concepts, such as interactions between goals and plans, while being relatively context-independent (Schank, 1982). For example, a thematic structure about "retaliation" may include knowledge that an actor A feels wronged by an action of B and then takes action meant to displease B. The particular actors and actions are not specified; such a structure can apply both to siblings tattling on each other and gangland killings. Thus, thematic knowledge structures can be applied cross-contextually, reflecting a similarity in the point of two episodes that may have few contextual features in common (Schank, 1982).

The form and content of such structures depends on, and determines, their function. TOPs help one understand new episodes similar to old ones, provide expectations about the episode, and let one predict what is to come. TAUs, a subset of TOPs, focus more specifically on knowledge about failures in planning situations. TAUs are based on interactions of goals and plans that are reflected in adages shared by a culture, especially those involving expectation and planning failures. They then contain planning information about problems that can occur, and how to avoid or solve them. Both TOPs and TAUs organize storage of individual episodes in memory, so that cases having the same theme are stored similarly in memory.

For these structures to be useful, they need to be accessed at an appropriate time to provide useful information when processing a new, similar case. Schank (1982) proposes that understanding a new situation requires, first, categorizing the input according to a structure in memory and second, retrieving an episode that is similar to the input. In general, to make use of any expectations and predictions a knowledge structure contains, the input must be matched to a structure similar to it.

The present study looks how much of the thematic pattern need be present in the input to result in a reminding. The "whole structure" argument says that the input and the retrieved episode must share *all* abstract features in a reminding; one cannot retrieve an episode based on only *some* of the input's abstract features. If the partial and the whole structure led to retrieval of the same exact episode, this would imply that the abstract features left out of the partial structure were irrelevant or coincidental. (Birnbaum, 1986). However, according to this argument, a change in the abstract relationships within the input will affect the likelihood of retrieving a certain instance, and a reminding appropriate to one input is likely to be inappropriate when elements of the input are changed.

However, if one must receive the whole thematic pattern in an input before a TOP or TAU could be activated, this would severely limit the usefulness of these knowledge structures. The expectations, predictions, and planning information they contain could not be used until the situation had played itself out. For example, activating the TAU "closing the barn door after the horse has gone" *after* one has made this planning error would not allow one to take advantage of that knowledge and avoid the failure. To be maximally useful, a TAU should be activated *before* the outcome of a planning situation; this would require accessing the TAU based on recognizing a partial thematic structure in the input.

The present study used stories based on TAU structures to test the whole structure hypothesis, using a reminding paradigm. Subjects read TAU-based stories in a study phase, and then in a test phase were presented with stories based on the same TAUs as in the study phase. Each TAU-based test story had two versions: a partial version with information that set up conditions for a planning failure but did not allow certain identification of a theme, and a full version that completely duplicated the partial but also included the outcome of the story, thereby providing a complete thematic pattern. If the partial structure stories led to remindings at a chance rate, this would support the total structure hypothesis (and cast some doubt on the psychological validity of TAUs as planning aids). On the other hand, if partial and full theme stories result in remindings at an equal rate, one could conclude that a whole structure is unnecessary (at least in some cases). Or, finally, the results could show that having the full structure is better than just having the partial, but having the partial results in reminding at least some of the time, providing predictive functionality.

Method

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

The materials consisted of a study packet, a test packet, and a distractor task. The main plot of each story featured a different abstract thematic pattern reflecting a planning failure, such as "don't count your chickens before they've hatched." In addition, each story contained different contextual details, such as taking place in a school setting or involving family interactions.

Twelve different stories were used in the test packets. Four stories were used as fillers, each having themes distinct from those featured in study packet stories and from each other. Each of the remaining eight test stories had the same theme as one of the stories from the study packet. That is, the theme of the study stories, such as "counting your chickens before they've hatched," was recreated in a test story, but in a different contextual setting. Therefore, each study story and its paired test story had only the thematic pattern in common. No other similarities in character or setting were included. Each test story had two versions: a partial version, which included the initial elements of the thematic pattern but not the conclusion or resolution of the story, and a full-length version, which completely duplicated the partial version but also provided the conclusion or outcome of the story. Each partial story provided enough information to set up conditions under which the planning failure reflected in its theme could occur. However, not enough information was included to allow certain identification of a theme. An example study story, partial test story, and full-length test story are shown in Table 1.

Table 1
Sample story materials illustrating experimental manipulations.

Study Story:
Dr. Popoff depended on his graduate student Mike's hard work in the lab, but he knew that Mike was unhappy with the research facilities available in the department. Mike had requested new equipment on several occasions, but Dr. Popoff always put off Mike's requests. One day, Dr. Popoff found out that Mike had been accepted to study at a rival university. Not wanting to lose a good student, Dr. Popoff hurriedly offered Mike lots of new research equipment. But by then, Mike had already decided to transfer.

Test Story - Partial Version
Melissa's aging grandmother live in a nursing home in a nearby town. Melissa and her grandmother had always been close and dearly enjoyed seeing each other. Lately, she knew, her grandmother's health had been worsening. Her grandmother often called from the home and asked Melissa to come and visit, but Melissa wondered when she would ever find time for a visit in her busy work schedule.

Test Story - Full-length Version
Melissa's aging grandmother lived in a nursing home in a nearby town. Melissa and her grandmother had always been close and dearly enjoyed seeing each other. Lately, she knew, her grandmother's health had been worsening. Her grandmother often called from the home and asked Melissa to come and visit, but Melissa wondered when she would ever find time for a visit in her busy work schedule. When Melissa discovered that her grandmother had become gravely ill, she finally decided to go and pay her grandmother a visit. When she got to the hospital, her grandmother was in the intensive care unit and could not receive any visitors.

Each subject received four partial test stories corresponding to four of the study stories, and four full-length test stories corresponding to the remaining four study stories. The pairing of study story to full-length or partial test story was counterbalanced across subjects (i.e., subjects received different groupings of partial stories, not always the same four). Full-length and partial stories were placed in randomized order for each subject.

Each of the 12 test stories was followed by instructions, taken from Gentner and Landers (1985) and two response spaces were provided to allow for multiple remindings for each test story. The distractor task was a packet of eight insight problems to be solved, unrelated to the themes in the stories.

Subjects received a study packet and were instructed to read through the stories and study them for five minutes, as they would be tested on them later. All subjects read all the stories at least once in this study period. After five minutes the study packets were collected and subjects were told to work through problem-solving packets. They were instructed to time themselves on each problem, using a digital clock provided.

The subjects were interrupted after ten minutes of work on this task. The subjects were then given the test packets. They were verbally told there were not answers to each reminding problem in the test packet so they were not to "wrack their brains" for answers, and if one didn't come to them relatively quickly they were to go on to the next page. They were also told to work forwards through the packet and not look back at completed pages.

After finishing this phase, free recall memory for the study stories was tested. Subjects were asked to write down, on the back of the test packet, 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 packet.

Results

Each page of the test packet was coded by story type (full theme, partial, or filler) and response type (match, intrusion, or null). Each subject received four of each story type in the test booklet. Mention of the corresponding story from the study packet counted as a match, whether it occurred in the first or second response space on the page. Filler stories by definition had no matches. An intrusion 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 target stories (false alarms). False alarms accounted for less than 1% of the responses. The same quantitative and qualitative coding used with subject data was used for materials data as well. The filler stories had a total of 54 observations; the full theme and partial stories averaged 27 observations each.

Subjects on average wrote down 4.4 responses as remindings and could recall a mean of 4.8 study stories at the end. The overlap of these two measures was high. The number of study stories available to each subject was the number of different stories that were reported as recalls or remindings (or both); this mean was 5.5.

The mean number of responses by story and response type are shown in Table 2. Full theme stories were matched significantly

more often than partial theme stories with subjects as a random factor, $t(53) = 3.80$, $p < .0001$, but this difference only showed a trend toward significance with materials as a random factor, $t(14) = 1.76$, $p < .101$.

Table 2
Proportion of responses by story and response type.

STORY TYPE	RESPONSE TYPE	
	match	intrusion
full theme	.48	.18
partial	.34	.13
filler	------	.12

To determine whether the number of responses to full themes and to partial themes differed from chance, these means were tested against estimates of chance performance. Subjects' reported thematic and partial matches were corrected by a factor estimating chance performance, 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 study. Both theme matches and partial matches were significantly different than this chance estimate, $t(54) = 10.07$ and $t(54) = 8.22$, respectively. A second estimate, looking at performance only on trials where the subject had the matching story available in memory, showed that both full themes and partial ones were matched at higher-than-chance level, $t(54) = 8.23$ and $t(54) = 7.644$, respectively.

The story types did not differ in proportion of intrusion responses when analyzed by either subjects or materials [$F(2,53) = 2.48$; $F(2,17) = 1.01$, respectively], but did differ in proportion of null responses under both analyses: for materials, $F(2,17) = 19.365$, $p < .0001$, and for subjects $F(2,53) = 113.75$, $p < .00001$. Filler stories evoked the fewest responses.

The test stories differed in length, in terms of number of words, as well as differing in type (as defined above), so analyses were done comparing long stories (full theme and filler) with short ones (partials). The number of responses, by story length, was significantly different when analyzed with subjects as a random factor, $t(53) = 2.72$, $p < .009$, with more responses being made to the shorter stories (all thematic) than to the long stories (half thematic and half not). A materials analysis of length was not significant.

Two coders rated the quality of the recall protocols, using a 5-point scale based on one used in Gentner and Landers (1985). The correlation of the coders' ratings, with agreements about null responses omitted, was .78. The ratings of the two coders were averaged to get a quality rating for each response. Number of words written was also calculated for each response; this measure was correlated with the quality measure at .71.

The average rating per response, across all story types, was 3.169. Mean quality ratings by story and response type are shown in Table 3.

Table 3
Average quality of responses by story and response type.

STORY TYPE	RESPONSE TYPE	
	match	intrusion
full theme	3.40	2.94
partial	3.17	2.64
filler	------	3.13

Between subjects t-tests of quality ratings were not significant for partial versus full theme matches, full theme matches versus filler intrusions, or partial matches versus filler intrusions. Analyzed by materials, the mean quality ratings of full theme versus partial matches showed a trend towards significance, $t(14) = 2.0$, $p < .065$. The two theme conditions did not differ in quality of intrusions when analyzed by materials, $t(14) = 1.45$, $p > .05$.

A within-subjects t-test of the quality of theme matches versus the quality of partial matches was significant, $t(38) = 2.55$, $p < .015$, with full theme responses being of higher quality. Further within-subjects t-tests of the quality of partial matches versus filler intrusions and versus partial intrusions were not possible, given the low number of subjects who had both protocols necessary for the comparisons.

Quality ratings were also analyzed by length of story. Analysis with subjects as a random factor showed a trend toward significance, $t(53) = 1.65$, $p < .105$, with long stories having responses of higher quality. This was also significant with materials as a random factor, $t(14) = 2.7$, $p < .017$.

Discussion

The results provide evidence that partial feature sets can elicit remindings based on thematic structures and can be used to access relevant cases during the planning process, rather than only after the outcome, and therefore the full pattern, is determined. Access is more likely when a whole thematic pattern is present, but partial patterns were also shown to be useful. The partial feature sets involved in the stories in this study were conditions that causally implied a certain outcome, and it may be that selection of other abstract "parts" of the pattern would not have the same effect. However, these conditions, and their relationship with the parts omitted, would be ones particularly relevant to planning situations.

The whole structure argument suggests that a partial pattern is unlikely to result in remindings because it does not completely specify a unique thematic pattern. For example, one could imagine similar set-up conditions for the adages "two heads are better than one" and "too many cooks spoil the broth." According to this argument, a change in conclusion would make the reminding inappropriate; the two adages are not analogous. However, in a planning situation getting reminded of either or both would be appropriate, as either one could provide planning information helpful in avoiding a bad outcome or making a good outcome more likely. The results show that

partial thematic patterns lead to access to instances in memory at better than chance levels.

The number of responses given was not related to the length of the test story presented. Rather, subjects were able to discriminate the thematically related stories from those not so related. This was clearly true with the full-length stories, although since no partial-length filler stories were provided it cannot be concluded with certainty that equal discrimination would occur among different types of partial stories.

The quality ratings of the protocols also show that subjects had access to the themes of the study stories. Both full-length and partial thematic stories had an average rating of over 3 (basic thematic elements present), and the observations consistently rated below 3 were mainly intrusions, suggesting that when a match was made, the test story provided better retrieval cues. The difference in quality between the responses to partial and full thematic patterns may result from the test stories having more complete reconstruction cues in the full than in the partial condition.

In summary, the results supported the hypothesis that people can use a partial set of thematic features to access related themes and cases stored in memory. The partial sets were less reliable for use in retrieval than full sets were, but they allowed some successful access. This is inconsistent with the whole structure argument, which says that reliable retrieval requires having a full thematic pattern evident in an exemplar. The results suggest that partial feature sets can serve a predictive function, allowing access to knowledge and expectations stored in thematic knowledge structures before and episode is completely determined, which in turn would help in avoiding potential planning failures. Future work will investigate whether there are particular partial sets of features that result in better access than other partial sets. For example, providing the initial conditions of a problem may lead to more reminding than when only a planning decision or outcome is provided. The functionality of the reminding constrains processing such that not all features will have equally valuable predictive ability to retrieve related information while there is still time to incorporate it into planning and problem solving.

References

Birnbaum, L. (1986). Integrated processing in planning. Doctoral Dissertation, Computer Science Department, Yale University.

Dyer, M. G. (1983). *In-depth understanding: A computer model of integrated processing for narrative comprehension.* Cambridge, MA: MIT Press.

Lehnert, W. (1980). Plot units and narrative summarization. *Cognitive Science 5:* 293-331.

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.

Schank, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people.* New York: Cambridge University Press.

Seifert, C. M., & Hammond, K. J. (1990). Intelligent Encoding of Cases. *Proceedings of the AAAI Spring Symposium Series on Case-Based Reasoning,* Palo Alto, California.

Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Human Learning and Memory, 12* (2), 220-231.