Analogy and Similarity: Determinants of Accessibility and Inferential Soundness

Mary Jo Rattermann
Dedre Gentner
Department of Psychology
University of Illinois at Urbana-Champaign
603 East Daniel St.
Champaign, Illinois 61820

Abstract

Analogy and similarity are widely agreed to be important in learning and reasoning. Yet people are often unable to recall an analogy which would be inferentially useful. This finding suggests that a closer examination of the similarity factors that promote retrieval is necessary. We approached this problem by investigating the role of relational commonalities (higher-order relations and first-order relations) and common object-descriptions in the accessiblity and inferential soundness of an analogy.

Subjects first read a large number of stories. One week later they were given a new set of stories to read. These new stories were designed to form matches which shared different combinations of object-descriptions, first-order relations and higher-order relations with the original stories. Subjects were asked to recall any stories from the original set that came to mind. Afterwards they rated the matches for subjective soundness and similarity.

The results of two experiments showed that subjects recalled the original stories that shared common object descriptions and first-order relations with the new stories. These results support the idea that similarity based access is enhanced by a combination of surface similarity and first-order relations. They also suggest that common higher-order relations play a smaller part in recall. In contrast, in both the soundness-rating and similarity-rating tasks subjects rated the pairs that shared higher-order relations higher than the pairs which shared surface similarity. This suggests that those aspects of similarity that govern recall are different than those aspects that govern similarity-ratings and soundness-ratings.

Analogy and similarity are widely agreed to be important in learning and reasoning. But recent evidence has shown that people are often unable to retreive analogies which, if retrieved, would be inferentially useful (Gick & Holyoak, 1980, D1983; Reed, Ernst & Banerji, 1974; Ross, 1984, in press). This research suggests a closer examination of how similarity promotes retrieval. Gentner and Landers (1984) approached this question

by comparing the retrievability and the subjective soundness of different kinds of similarity matches. Because our studies build on their method, we begin by describing this study and the theoretical issues that led up to it.

The theoretical framework for the research is Gentner's (1980, 1983) structure-mappping theory. The basic intuition of structure-mapping theory is that an analogy is a mapping of knowledge from one domain (the base) into another (the target) which conveys that a system of relations that holds among the base objects also holds among the target objects. Thus an analogy is a way of noticing relational commonalities independently of the objects in which those relations are embedded. In interpreting an analogy, people seek to put the objects of the base in 1-to-1 correspondence with the objects of the target so as to obtain maximum structural match. The corresponding objects in the base and target don't have to resemble each other at all; object correspondences are determined by roles in the matching relational structures. Central to the mapping process is the principle of **systematicity**: people prefer to map systems of predicates that contain higher-order relations with inferential import, rather than to map isolated predicates. The systematicity principle is a structural expression of our tacit preference for coherence and deductive power in interpreting analogy.

Besides analogy, other kinds of similarity matches can be distinguished in this framework, according to whether the match is one of relational structure, object descriptions, or both. Recall that *analogies* discard object descriptions and map relational structure. *Mere-appearance* matches are the opposite: they map aspects of object descriptions and discard relational structure. *Literal-similarity* matches map both relational structure and object-descriptions.

It is helpful for this discussion to decompose analogical reasoning into access and mapping-plus-inference. *Access* is the process of retrieving a base situation in memory, given a target situation that the learner is currently considering. *Mapping* occurs after a base situation has been accessed from memory. In mapping, the predicates of the base are matched with predicates of the target according to the rules given above. Further predicates are carried across and inferences may be drawn. According to structure-mapping, the subjective soundness of a possible analogy depends on the degree to which a systematic relational match can be found.

Against this background, the Gentner and Landers (1985) experiment had a two-fold purpose: (1) it tested the prediction that shared systematic structure determines the subjective soundness of a match; and (2) it asked whether the accessibility of analogy and other kinds of similarity matches mirrors their subjective soundness. The study was designed to create a situation resembling natural long-term memory access. Subjects were first given about 30 stories to read and remember. One week later, they read a new set of stories and reported any cases in

which they were reminded of any of the original stories. The stories were carefully designed to embody three different kinds of similarity matches: *mere appearance*, *true analogy* and *false analogy*. In *mere-appearance (MA)* matches, the base and target shared object descriptions (e.g. a hawk in the base story vs. an eagle in the target) and first-order relations (e.g. shoot at [x,y] vs. fire at [x',y']). In *true-analogy (TA)* matches, the base and target shared first-order relations and higher-order relations (i.e., relations between relations, such as CAUSE [S(x,y), R(y,z)] vs. CAUSE [S'(x',y'), R'(y',z')] and other constraining relations) but not object-descriptions. In *false-analogy (FA)* matches only the first-order relations matched. (See Table 1, below, for examples.) Note that in all three cases the base and target shared first-order relations; the three similarity conditions differed in which, if any, other commonalities also existed.

Gentner and Landers found that the proportion recalled for the mere-appearance matches was greater than that of the true-analogy matches, which in turn was greater than that of the false-analogy matches. These results suggest that access to memory is heavily influenced by surface commonalities. In contrast, when the same subjects rated the inferential soundness of the similarity matches, the true-analogy matches were rated as the most sound, while the false-analogy and the mere-appearance matches were rated significantly less sound. As predicted, subjective soundness depended on common systematic structure. This raises an interesting disassociation: it seems that the kinds of similarity matches that subjects consider most sound are not the matches that most strongly promote access. Common surface information, such as shared object-descriptions, may have a disproportionate affect on accessibility.

The Gentner and Landers study revealed some interesting points about analogical access, but also left a great many questions unanswered. While the study suggested that surface similarity plays a large role in remindings, it also suggested a role for higher-order relations in remindings, since true-analogy matches were better accessed that false-analogy matches. However, these results do not tell us how (or whether) surface attributes and higher-order relations combine to promote access, nor do they tell us which aspects of surface similarity are most effective in promoting retrieval.

In Experiment 1 we addressed the first of these questions. To do this we replicated the Gentner and Landers study, adding another match type, that of literal similarity. A literal similarity match has commonalities at all levels -- object-attributes, first-order relations and higher-order relations. This meant that the set of matches formed a 2 X 2 design as shown in Figure 1 (below). Based on the results of the Gentner and Landers study we expected that literal-similarity matches would be recalled well in the reminding task. However, it was not clear how much the addition of higher-order commonalities would affect the results. They might contribute very little beyond the

effects of surface commonalities, in which case recall of the literal-similarity matches should not differ significantly from that of the mere-appearance matches. Alternately, it could be that having commonalities at all levels would lead to far greater accessibility for literal-similarity matches than would be predicted by combining the two separate effects.

One problem with the Gentner and Landers study is that it leaves open an alternative interpretation of the results: namely, that the retrievability ordering among the three types of matches was simply a function of their overall similarity. That is, it could have been the case that the mere-appearance matches were more similar to one another than the true-analogy matches, and the true-analogies more similar than the false-analogies. In this case, there would be no reason to invoke a special role for surface attributes in similarity-based access. To address this possibility, we added a similarity-rating task in order to test whether retrievability could simply be predicted from similarity ratings.

To summarize, in this study three measures were obtained for each pair of scenarios: (1) the accessibility of the first story given the second, i.e., how well the second story served as a cue for the first; (2) the inferential soundness of the analogy between the stories, as rated by the subjects; and (3) the degree of similarity between the two stories, as rated by the subjects.

Experiment 1

Method

Subjects

The subjects were 36 undergraduates who received class credit for participation in this experiment. Due to experimenter error, 18 paid subjects were used to rerun one cell in the soundness-rating task and one cell in the similarity-rating task.

Materials

There were 20 story sets, each consisting of a base story plus four different target stories designed to embody different kinds of similarity match: *literal similarity (LS)*, *mere appearance (MA)*, *true analogy (TA)*, and *false analogy (FA)*. Figure 1 shows the design of the match types; Table 1 shows example stories. In addition to the 20 story sets, there were 12

	HIGHER-ORDER RELATIONS	
	SHARED	NOT SHARED
OBJECT ATTRIBUTES SHARED	Literal Similarity	Mere Appearance
OBJECT ATTRIBUTES NOT SHARED	True Analogy	False Analogy

Note: First-order relations are roughly constant.

Figure 1: Design of Match Types

Table 1: Example Stories

BASE Story

Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so greatful that he pledged never to shoot at a hawk again. He went off and shot a deer instead.

Literal similarity

Once there was an eagle named Zerdia who nested on a rocky cliff. One day she saw a sportsman coming with a crossbow and some bolts that had no feathers. The sportsman attacked but the bolts missed. Zerdia realized that the sportsman wanted her tailfeathers so she flew down and donated a few of her tailfeathers to the sportsman. The sportsman was pleased. He promised never to attack eagles again.

True Analogy

Once there was a small country called Zerdia that learned to make the world's smartest computer.

One day Zerdia was attacked by it's warlike neighbor, Gagrach. But the missiles were badly aimed and the attack failed. The Zerdian government realized that Gagrach wanted Zerdian computers so it offered to sell some of it's computers to the country. The government of Gagrach was very pleased. It promised never to attack Zerdia again.

Mere Appearance

Once there was an eagle named Zerdia who donated a few of her tailfeathers to a sportsman so he would promise never to attack eagles.

One day Zerdia was nesting high on a rocky cliff when she saw the sportsman coming with a crossbow. Zerdia flew down to meet the man, but he attacked and felled her with a single bolt. As she fluttered to the ground Zerdia realized that the bolt had her own tailfeathers on it.

False Analogy

Once there was a small country called Zerdia that learned to make the world's smartest computer. Zerdia sold one of its supercomputers to its neighbor, Gagrach, so Gagrach promised never to attack Zerdia.

But one day Zerdia was overwhelmed by a surprise attack from Gagrach. As it capitulated the crippled government of Zerdia realized that the attacker's missiles had been guided by Zerdian supercomputers.

additional stories which Were used as fillers. Each subject received 20 base stories: five of each of the four match types (LS, MA, TA, and FA). Each subject saw one and only one target story for each base story. The subjects were divided into four groups in order to counterbalance the assignment of stories to match types.

Procedure

Reminding Task. In the first session subjects were told to read and remember 20 base stories and 12 filler stories. One week later, they returned for the reminding session. They were given booklets of 20 target stories, each of which matched one and only one base story. For each target story, subjects were told to notice if they were reminded of any story from the previous session, and to write down any such stories in as much detail as possible.

Soudness-rating Task. The soundness-rating task was given to subjects after they had completed the reminding task. Subjects were shown the same 20 pairs of stories they had received in the reminding task, with each pair consisting of a base story and a matching target story. They rated each pair for the soundness of the match between the two stories. A sound match was described as one in which the two situations match well enough so that inferences in one would be likely to carry over to the other. This description was placed in the context of what would make a good argument. Subjects used a 1-5 scale, where 5 = highly sound and 1 = spurious.

Similarity-rating Task. Following the soundness-rating task subjects were again given the same pairs of stories and were asked to rate the pairs on their overall similarity. [1] No explicit definition of similarity was given; we simply allowed subjects to use their own intuitions.

Scoring the Reminding Task. To score the reminding task we had to judge whether the subjects had indeed successfully retrieved the original stories (as opposed to just guessing, for example). The recalls were scored in three ways, but we focus here on one method only. Two judges scored on a 1-5 scale how well subjects recalled each base story by comparing the subject's recall with the correct base story. Then we computed for each match type the proportion of stories for which a rating of 2 or better -- indicating that the subject's description had clearly mentioned at least a few elements of the base -- had been assigned. This *flat match* score was designed to capture whether any genuine retrieval of the correct base story had occurred, without concern for whether the recall was of high quality.

---

1. As a check on whether the prior soundness-rating task had any influence on the subsequent similarity-rating task a later group of subjects was run with the reverse order of the soundness and similarity rating tasks. The results were unchanged.

Results

Reminding task

As shown in Figure 2a, literal-similarity and mere-appearance matches led to significantly more reminding than the true-analogy or false-analogy matches.[2]

These results replicated the pattern found by Gentner and Landers in that the mere-appearance matches produced more remindings than true-analogy matches. These results support the idea that similarity-based accesss is enhanced by common surface similarity, and that common relational structure plays a smaller part in recall.

Soundness-rating task

Subjects rated the literal-similarity and the true-analogy matches as significantly more sound than the false-analogy and mere-appearance matches. (See Figure 2b.)

Again, these findings replicated those of Gentner and Landers in that true-analogy matches are considered significantly more sound than false-analogy matches and mere-appearance matches. Further, the fact that literal-similarity and true-analogy -- the two match types that utilize common higher-order structure -- were considered significanty more sound than the other two match types provides evidence for the prediction of the structure-mapping theory that soundness is governed by common relational structure.

Similarity-rating Task

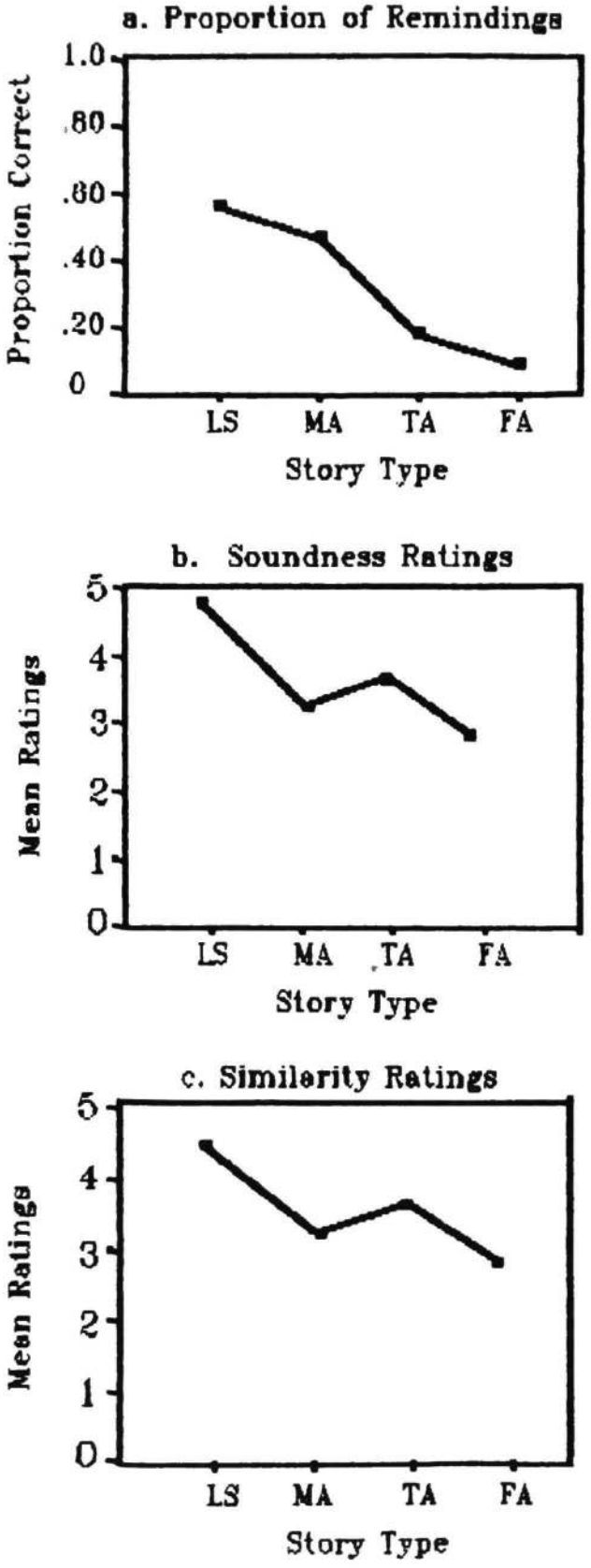
Subjects rated the literal-similarity matches as significantly more similar than the true-analogy matches, and the literal-similarity and true-analogy matches were both significantly more similar than the mere-appearance and false-analogy matches. However, the mere-appearance matches and the false-analogy matches were not significantly different from each other. (See Figure 2c.)

The similarity results showed an interesting trend: the pattern of the similarity ratings mirrored that of the soundness data. Specifically, subjects rated the pairs that shared higher-order relations (the literal-similarity matches and the true-analogy matches) as more similar than the pairs which shared surface attributes (the mere-appearance matches). This suggests that in making similarity ratings subjects judge higher-order relations to be more important than surface attributes. However, it should be noted that literal-similarity matches received higher similarity ratings than true-analogy matches, showing that

---

2. One way analyses of variance were performed on the data sets in Experiment 1. In each case the analysis revealed a main effect of Match type, $p < .001$. Post hoc analyses (Tukey's, alpha = .05) were used to determine the significant differences between the means. For brevity we will omit the statistics for the remainder of the results and simply report significant differences.

Figure 2
Results from Experiment 1



(unlike soundness) similarity is also sensitive to surface features. The other important implication of the similarity findings is that, since they do not mirror the accessibility ordering, it is improbable that the accessibility results are simply an artifact of differences in similarity across match types. On the contrary, the order of accessibility match types is quite different from the order of similarity ratings.

The most important finding here lies in the comparison between the results of the recall task and the results of the similarity-rating and soundness-rating tasks. It seems that different aspects of similarity govern these different processes. In similarity-based recall, it is common surface features such as object descriptions that matter most, while in judging the soundness or similarity of two stories it is common relational structure that matters most.

Experiment 2

Experiment 2 was designed to examine more closely the issue of "surface commonalities". In the studies so far, the mere-appearance matches share both object-descriptions and first-order relations. Here we asked which aspects of mere-appearance matches led to their accessibility; and in particular, whether object-descriptions alone could promote access. In this study we created an new variant of mere-appearance matches by removing all the first-order relational commonalities from the first set, leaving the common object-descriptions (as exemplified below).

> There once was a sportsman who loved to hunt. He liked to have the animals he caught stuffed and mounted. His pride and joy was an eagle he had killed with just a crossbow and a bolt. He had been hiding in the top of an elm tree when he shot her.

We then ran the study again, pitting these new matches that shared only object-level attributes (called mere-appearance-attribute-only, or MAAO) against the mere-appearance matches that shared both object attributes and first-order relations (called mere-appearance-first-order, or MAF) against true-analogy matches and literal-similarity matches.

Method

Subjects

The subjects were 52 undergraduates who were fulfilling a course requirement.

Materials and Procedure

Subjects were divided into two groups. Each group received 14 true-analogy matches, and 14 mere-appearance matches, with half the subjects receiving mere-appearance-first-order versions and half receiving mere-appearance-attribute-only versions. In addition, each group received 6 literal-similarity matches to anchor their responses. The procedure was as in

Experiment 1. In the first session, the subjects read 20 base stories and 12 filler stories. One week later, the subjects were given 20 new stories: 7 TA matches, 7 of either one of the mere-appearance matches (either MAAO or MAF) and 6 LS matches, followed by a soundness-rating task and a similarity-rating task. Again, as in Experiment 1, each subject saw only one target story for each original base story.

Results

Reminding task

The results, shown in Figure 3a, were rather dramatic. First, LS matches and MAF matches were recalled significantly better than MAAO matches, which were recalled significantly better than TA matches. [3]

These results show that it is not just common object-descriptions that lead to similarity based recall; rather the biggest gain in accessibility seems to occur for some combination of common object-descriptions and common first-order relations. In fact, the match type that possessed common object-descriptions with the base, mere-appearance attributes-only, was recalled quite poorly. What is particularly striking about these results, however, is that TA matches were even less accessible than MAAO matches. As discussed below, this ordering is again in sharp contrast to the subject's own opinions concerning the inferential soundness and similarity of the matches. Again we find that the ordering of accessibility is different from the ordering of soundness.

Soundness-rating task

As shown in Figure 3b, the order of soundness ratings was LS matches followed by MAF matches followed by MAAO matches. (All of these differences were significant.)

These findings again confirm that the most important determinant of inferential soundness in subjects' judgements is common systematic relational structure (which is present only in the LS and TA matches). However, in this study, unlike the prior two, subjects considered LS matches slightly more sound than TA matches, indicating that for these subjects common object descriptions also contribute to soundness (this being the only difference between LS and TA). Not surprisingly, the MAAO matches, which share only the cast of characters, were considered least sound.

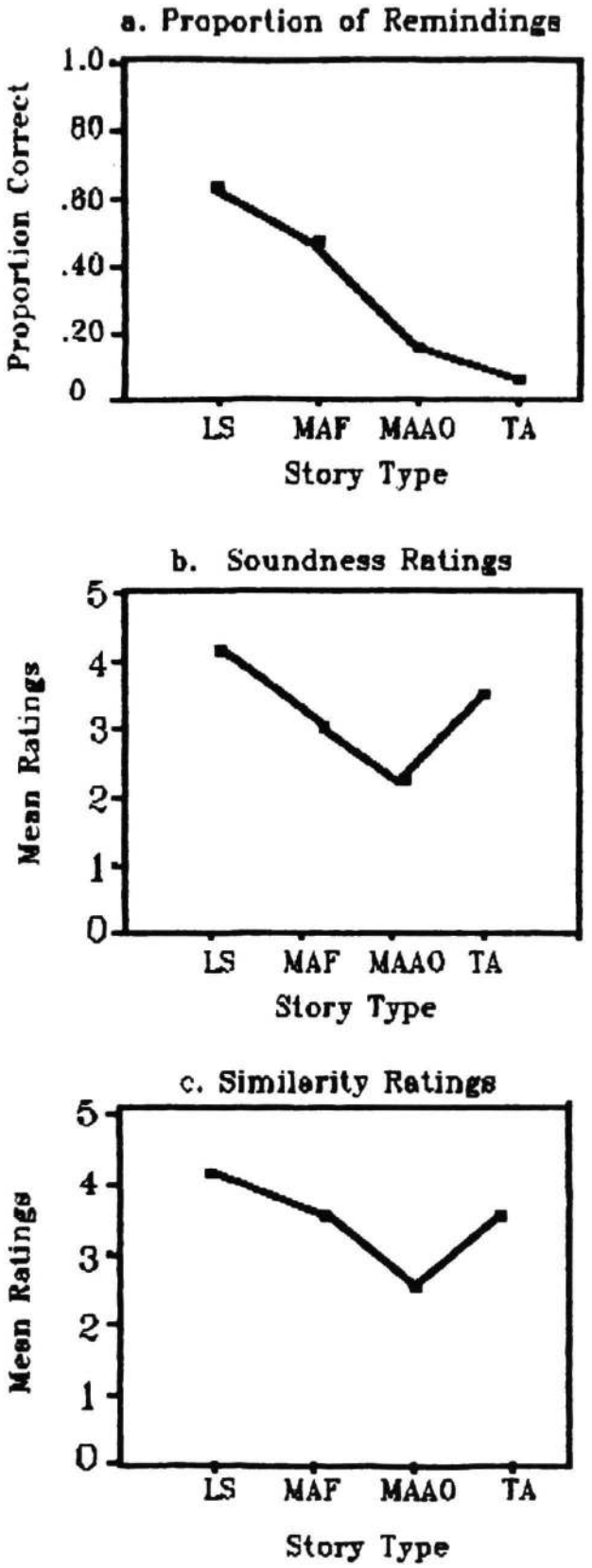
Similarity-rating task

Again, the similarity ranking are close to the soundness rankings: the order of similarity ratings is LS, TA and MAF, and then MAAO. (All differences are significant *except* that between TA and MAF.) Quite reasonably, subjects consider matches with

---

3. Again to be brief we will only report significant results in the body of the paper. Significance was determined by Welch-Aspin t-tests, alpha<.05.

Figure 3
Results from Experiment 2



three levels of commonality (LS matches) to be the most similar, and matches with only one level of commonality (MAAO matches) to be least similar. The other two match types each share two levels (TA matches share higher-order and first-order relations, while MAF matches share first-order relations and object descriptions) and subjects here considered them roughly equally similar. (This differs slightly from Experiment 1, in which TA matches were rated as more similar than MAF matches.)

The most important point about the similarity ratings, is that they again demonstrate that the accessibility results are not due to perceived similarity. As before, the pattern of similarity is very similar to the patterm of soundness and both are quite different from the order of the accessibility.

Discussion

The most important finding here is the dissociation between the kinds of similarity that people think are inferentially reliable and the kinds of similarity that readily enable memory access. This disparity can be seen most strikingly in the comparison between true analogies and mere-appearance matches. True-analogy matches are consistently felt to be more sound than mere-appearance matches; yet mere-appearance matches are consistently better at promoting retrieval. These accessibility results cannot be attributed to differences in similarity, for the true-analogy matches are consistently felt to be as similar or more similar than the mere-appearance matches.

Thus we are left with the disturbing findings that the pattern of accessibility is very different from the patterns of subjective inferential soundness. These results are compatible with the finding that people in problem solving tasks often fail to retrieve prior analogical problems which, if retrieved, would help them solve the current problem (Gick and Holyoak, 1980, 1983; Reed, Ernst and Banerji, 1974, Ross, 1984, in press)

Such results are problematic for models of memory that assume heavy reliance on causal indexing (e.g. Schank, 1982). On the contrary, it appears that people tend to have a fairly surface-oriented default indexing scheme that emphasizes object and first-order relations.

A natural question at this point is why humans should have such as seemingly arbitrary method of memory indexing and retrieval. One part of the answer may lie in the performance of the literal similarity matches. These matches are high on every measure: they are highly accessible, they are considered extremely sound and extremely similar. One speculation is that our memory indexing is geared toward literal-similarity matches: cases in which things resemble each other on the surface and also share deeper relational structure. In many areas the strategy may work fairly well; often thing that look alike are fundamentally the same. But in cases where appearance is not a good predictor of underlying relational structure, our memory systems may play us false.

References

Gentner, D. (1980). The structure of analogical models in science (BBN Rpt. No. 4451). Cambridge, MA: Bolt Beranek and Newman Inc.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7(2), 155-170.

Gentner, D., & Landers, R. (1985, November). Analogical reminding: A good match is hard to find. In Proceedings of the International Conference on Systems, Man and Cybernetics (pp.607-613). 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), 1-38.

Reed, S. K., Ernst, G. A., & Banerji, R. (1974). The role of analogy in Transfer between similar problem states. Cognitive Psychology, 6, 436-450.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. Cognitive Psychology, 16, 371-416.

Ross, B. H. (in press). Remindings in learning: Objects and tools. To appear in S. Vosniadou & A. Ortony (Eds.). Similarity and analogical reasoning.

Schank, R. C. (1982). Dynamic memory. New York: Cambridge University Press.