A Model of the Role of Expertise in Analog Retrieval

L. Karl Branting
Department of Computer Science
University of Wyoming
Laramie, Wyoming 82071-3682
karl@eolus.uwyo.edu

Abstract

This paper presents a model of the use of expert knowledge to improve accuracy of analog retrieval. This model, match refinement by structural difference links (MRSDL), is based upon the assumption that expertise in domains requiring analogical reasoning consists in part of knowledge of the structural similarities and differences between some pairs of the source analogs. In an empirical evaluation on four data sets, MRSDL consistently retrieved the most similar or nearly most similar source analog. Achieving comparable accuracy on these data sets with a two-stage retrieval technique such as MAC/FAC would require exhaustive matching with more than half of the source analogs. The evaluation also showed that parallel competitive matching is often substantially faster than exhaustive matching or MRSDL.

Similarity in Analogical Reasoning

The terms "reasoning by analogy" and "case-based reasoning" subsume a variety of different problem-solving and learning activities. Common to all these activities, however, is attributing conclusions to a new situation based on its relevant similarity to some previous situation to which the same conclusions applied.

There is a consensus among researchers in analogical reasoning that *structural consistency* is a central component of similarity for the purposes of analogical reasoning (Winston, 1980; Gentner, 1983; Falkenhainer et al., 1989; Holyoak and Thagard, 1989). Two analogs are structurally consistent if objects in the two analogs can be placed into correspondence so that relations also correspond. This correspondence is generally modeled as a mapping from the objects in one analog (the *source*) to those in another (the *target*).

A number of factors have been identified that may influence the process of constructing a mapping from a source to a target analog. Holyoak *et al.* (Holyoak and Thagard, 1989) have stressed the role of *semantic similarity*, preference for mappings that put semantically similar objects and relations into correspondence, and *pragmatic centrality*, preference for mappings that are directly related to problem solving goals, in constraining the mapping process. Genter has emphasized *systematicity*, preference for mappings between "higher-order" relations, *i.e.*, those that take propositions as arguments, over first-order relations, *i.e.*, those that take objects as arguments. Other research, *e.g.*, (Faries and Reiser, 1990) and (Branting and Porter, 1991), has studied the effect of *elaboration* of the target analog, that is, inferring facts not explicit in the target analog. Finally, (Branting, 1991) and (Branting and Porter, 1991) illustrated use of general domain theory to reformulate a problem in a manner that can lead to improved structural consistency with its most similar analog. Following (Holyoak and Thagard, 1989), semantic similarity, pragmatic centrality, systematicity, target elaboration, and problem reformulation will be collectively referred to as *constraints* on the mapping process.

Methods for Analog Retrieval

The task of analog retrieval is to determine the potential source analog in memory that shares the greatest structural consistency with a target analog, or *probe*, under a given set of mapping constraints. The simplest approach to analog retrieval is exhaustive matching between a target analog and all potential source analogs in memory. However, exhaustive matching is psychologically implausible and computationally intractable for large knowledge bases.

Implemented alternatives to exhaustive matching include ARCS (analog retrieval by constraint satisfaction) (Thagard et al., 1990) and MAC/FAC (many are called but few are chosen) (Gentner and Forbus, 1991). Given a target probe, ARCS first finds a set of candidate source analogs that "in some degree" share semantic similarities with the probe. For each candidate analog, ARCS constructs a constraint network. A connectionist relaxation algorithm is then used to settle into a state that indicates the relative correspondence of the various stored structures to the probe under the given constraints. MAC/FAC is also two-stage model. A computationally inexpensive measure of surface similarity is used to retrieve an initial set of candidates. Exhaustive matching is then used to determine which of the candidates is structurally most similar to the probe.

ARCS and MAC/FAC both successfully account for the widely observed phenomenon that surface (*i.e.*, semantic) similarity is a stronger predictor of memory access in novices than structural consistency (Ross, 1989; Gentner, 1989) (although structural consistency is also a predictor of retrieval (Wharton et al., 1991)). There is reason to question, however, whether these approaches to retrieval are equally successful at modeling analog retrieval by experts. There is empirical evidence that experts are better than novices at using structurally similar analogs and are less prone to use analogs with misleading surface similarities (Novick, 1988). The hallmark of expertise in many fields is the ability to find the structurally most similar analog irrespective of surface differences. In legal reasoning, for example, the legal precedent most relevant to a given case may have very different facts. Skillful attorneys are adept at finding such precedents.

Modeling analogical retrieval in experts therefore requires showing how the most similar (or nearly most similar) source analog can be found without exhaustive search of memory. The difficulty of two-stage retrieval methods such as ARCS and MAC/FAC is in determining the size of the set of initial candidates. If the initial candidate set is too small, then the most similar analog may not be found. If the initial candidate set is too large, then exhaustive search of the candidate set will not be significantly less expensive than searching the entire library of analogs. When surface similarity is unreliable, a sufficiently poor choice of candidate set size can conceivably lead to the worst of both worlds: exhaustive search of a significant portion of the analog library that nevertheless fails to retrieve the most relevant analog. Improving upon the two-stage retrieval models requires showing how expert knowledge can improve retrieval accuracy.

One form of knowledge that experts can be expected to have and novices lack is knowledge of the structural similarities and differences between at least some pairs of the analogs in memory. Suppose, for example, that a law student is asked to analyze a hypothetical *H1*, and the student recalls a superficially similar precedent *P1*. Suppose that the student is then told that the controlling precedent is instead *P2* because of the greater structural similarities between *H1* and *P2*. To profit from this lesson, the student must understand the structural differences both between *H1* and *P1* and between *H1* and *P2* in order to appreciate that the former are greater than the latter. Perforce, the student must also understand the structural differences between *P1* and *P2* that led to the differences in their degree of structural similarity to *H1*. If on a later occasion the student encounters hypothetical *H2* that is superficially similar but structurally dissimilar to *P1*, the student can use knowledge of the structural differences between *P1* and *P2* to recover from the spurious match to *P1* and find the structurally most similar precedent *P2*.

The next section describes an algorithm that uses preexisting knowledge of structural differences among source analogs to recover from spurious matches.

Match Refinement by Structural Difference Links

In match refinement by structural difference links (MRSDL), an initial candidate is selected on the basis of surface semantic similarity. Precomputed information on the structural relations among analogs is then used to refine the match. Specifically, if the structural differences between an analog A_{cur} and a probe P have been determined, difference links containing precomputed information about the structural differences between A_{cur} and alternative analogs $A_1 \ldots A_n$ can be used to estimate inexpensively the similarity between P and each A_i. The idea behind this approach is that A_i is a better match to P than A_{cur} to the extent that A_{cur} differs from A_i and P in the same way. However, to the extent that A_i has additional differences from P, the match between A_i and P is worse. The most promising A_i is the case for which the differences with A_{cur} shared by A_i and P are greatest and the additional differences between A_i and P are least.

Consider a simple example involving the following brief narratives, represented in figure 1:

- **Probe.** John gave flowers to Mary because he likes her.
- **Analog-1.** Jimmy likes Billy because Billy gave him a snake.
- **Analog-2.** Bob gave flowers to Sally because he likes Sally's mother, Jane.

The highest degree of surface semantic similarity is between the probe and Analog-1: the probe and Analog-1 have identical relations, whereas Analog-2 has a relation, *mother*, not found in the probe. An initial retrieval based on surface semantic similarity would therefore favor Analog-1. However, Analog-1 differs structurally from the probe. The mapping that maximizes the structural congruence between the probe and Analog-1, Analog-1⇒Probe, is the following:

Billy→John
Jimmy→Mary
like2→like1

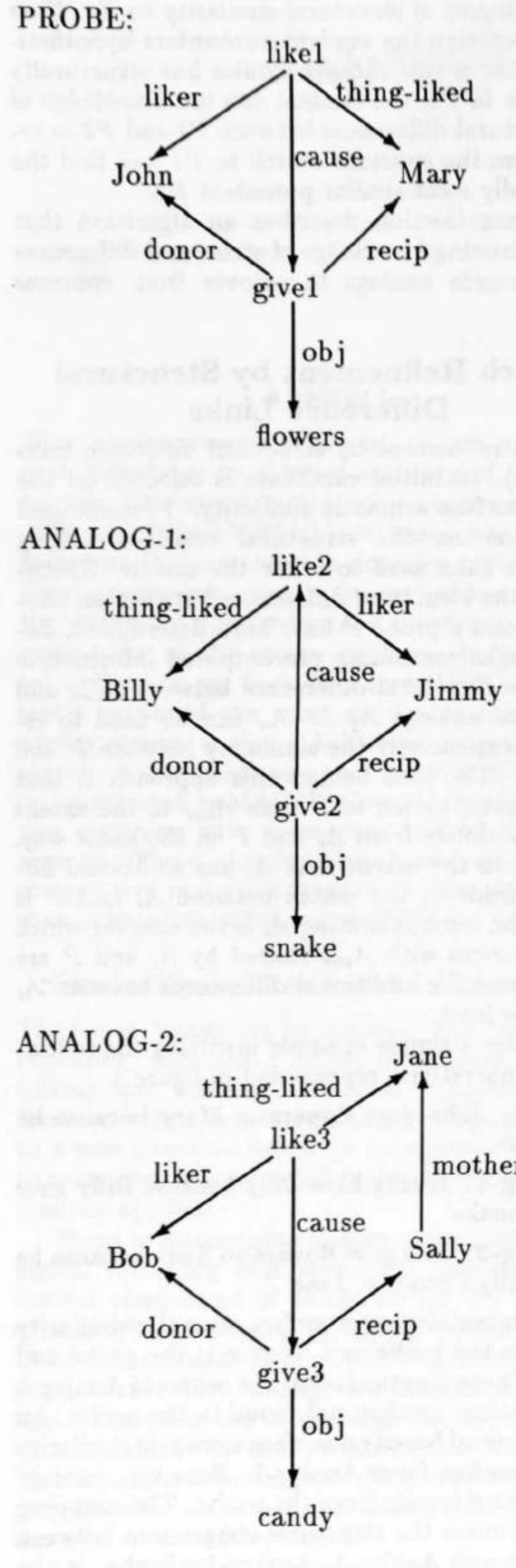

Figure 1: A probe and two analogs.

give2→give1
snake→flowers

Under this mapping, the following propositions in Analog-1 have no corresponding propositions in the probe:

(thing-liked like2 Billy)
(liker like2 Jimmy)
(cause give2 like2)

These unmatched propositions constitute a *difference* denoted Analog-1 − Dom(Analog-1⇒Probe) (where Dom(Analog-1⇒Probe) is the set of propositions having an image under Analog-1⇒Probe).

Assume that the following structural information concerning Analog-1 and Analog-2 has been precomputed:

- Analog-1⇒Analog-2, the mapping that maximizes the structural congruence between Analog-1 and Analog-2, and Dom(Analog-1⇒Analog-2), the Analog-1 propositions that have an image in Analog-2 under Analog-1⇒Analog-2.
- Analog-2⇒Analog-1, the mapping that maximizes the structural congruence between Analog-2 and Analog-1, and Dom(Analog-2⇒Analog-1), the Analog-2 propositions that have an image in Analog-1 under Analog-2⇒Analog-1.

Using this information, the number of propositions of Analog-2 that would have no image in the probe under the best mapping from Analog-2 to the probe can be estimated by the sum of the number of Analog-1 propositions having an image in Analog-2 but no image in the probe, *i.e.*, |Dom(Analog-1⇒Analog-2) − Dom(Analog-1⇒Probe)|, which in this case is zero, and the number of Analog-2 propositions that have no image in P under the composition of Analog-2⇒Analog-1 and Analog-1⇒Probe (*i.e.*, |Analog-2 − Dom(Analog-2⇒Analog-1 ∘ Analog-1⇒Probe)|. In this case, Analog-2 − Dom(Analog-2⇒Analog-1 ∘ Analog-1 ⇒ Probe) is the following:

(thing-liked like3 Jane)
(mother Sally Jane)

This is fewer than the three propositions in Analog-1 − Dom(Analog-1⇒Probe), so Analog-2 is a closer match to the probe than Analog-1.

The full algorithm for match refinement is as follows:

Given:

- P, a probe (*i.e.*, target analog)
- A_{cur}, the source analog that is currently the best match to P
- A_{cur} − Dom(A_{cur}⇒P), the propositions of A_{cur} that have no image in P under A_{cur}⇒P, the best mapping from A_{cur} to P

- Precomputed difference links between A_{cur} and cases $A_1 \ldots A_n$ containing the following information for each A_i:

  $A_{cur} \Rightarrow A_i$, the best mapping from A_{cur} to A_i, and $\mathrm{Dom}(A_{cur} \Rightarrow A_i)$, the propositions of A_{cur} that have an image in A_i under $A_{cur} \Rightarrow A_i$

  $A_i \Rightarrow A_{cur}$, the best mapping from A_i to A_{cur}, and $\mathrm{Dom}(A_i \Rightarrow A_{cur})$, the propositions of A_i that have an image in A_{cur} under $A_i \Rightarrow A_{cur}$

Do:

1. Select the A_i for which $|A_i - \mathrm{Dom}(A_i \Rightarrow P)|/|A_i|$, the proportion of propositions unmatched under the best mapping from A_i to P, is estimated to be least, where $|A_i - \mathrm{Dom}(A_i \Rightarrow P)|$ is estimated by the cardinality of the following set:

   $\mathrm{Dom}(A_{cur} \Rightarrow A_i) - \mathrm{Dom}(A_{cur} \Rightarrow P) \bigcup A_i - \mathrm{Dom}(A_i \Rightarrow A_{cur} \circ A_{cur} \Rightarrow P)$
2. Calculate the actual value of $A_i - \mathrm{Dom}(A_i \Rightarrow P)$
3. If $|A_{cur} - \mathrm{Dom}(A_{cur} \Rightarrow P)|/|A_{cur}| < |A_i - \mathrm{Dom}(A_i \Rightarrow P)|/|A_i|$, then P matches A_{cur} better than any of the A_i's, so return A_{cur}. Otherwise, call the procedure again with A_i as the current best match.

As illustrated in (Branting, 1991), the composition of two best-mappings may fail to be itself a best mapping. Under these circumstances the algorithm may either over- or underestimate the true degree of structural difference between an analog A_i and a probe. As a result, difference-link refinement is a heuristic procedure.[1]

Comparison of MRSDL with Other Retrieval Techniques

To determine whether MRSDL represents an effective model of the use of expert knowledge in analog retrieval, a comparative evaluation was performed in which MRSDL was compared to three other retrieval techniques. The first alternative retrieval technique was exhaustive matching. The second technique was Best-First Incremental Matching (BFIM) (Branting, 1991). BFIM consists of best-first search of the space of partial mappings between each analog in memory and the probe. BFIM resembles ARCS in that it is a form of parallel matching involving competition among analogs (although in BFIM this competition doesn't consist of inhibition between competing match hypotheses, but merely of directing computational resources to the most promising match). The third technique was surface semantic retrieval. Degree of surface match was determined by the proportion of relations occurring in an analog that also occurred in the probe.[2]

These techniques were compared on four sets of analogs. The first two consisted of 100 fables and 26 plays (25 Shakespearean plays and West Side Story), generously provided by Paul Thagard, consisting of approximately 21 propositions per play and 55 propositions per play. The remaining two sets of analogs, taken from the worker's compensation law knowledge base of GREBE (Branting, 1991), consisted of 11 precedents of employment activities (averaging 29 propositions per case) and 10 precedents of near-miss noninstances of employment activities (averaging 30 propositions per case).

In each retrieval trial, the fables and plays were randomly divided into 5 or 3 (respectively) approximately equal partitions. Each analog of each partition was then used as a probe with the cases of the remaining partitions as analogs. Thus, each retrieval of each fable was tested using 80 other fables as analogs, and retrieval of each play was tested using 17 or 18 other plays as analogs. A set of 21 worker's compensation hypotheticals (averaging 89 propositions per case) were used to test retrieval of the instance and noninstance precedents of employment activities.

Before MRSDL could be run on each collection of analogs, some set of difference links had to be installed among them. The behavior of MRSDL depends heavily upon the configuration of difference links among analogs (Branting, 1991). For example, if there is no sequence of difference links connecting an initial surface match with the closest analog, then clearly no series of match refinements can retrieve the closest analog.

In this experiment, no effort was made to achieve an optimal configuration. Instead, a configuration of analogs connected by difference links was incrementally built up in a manner consistent with the scenario presented at the end of section 2: Each configuration was initialized with a single randomly selected analog. The remaining analogs were added in random order. For each new analog A, a difference link was installed between A and the superficially most similar analog, $SS(A)$. Exhaustive search was then used to determine the analog structurally most similar analog, $Ex(A)$. If $Ex(A)$ and

[1] To compensate for the possible inaccuracy of the estimate of the degree of structural difference between an analog A_i and a probe, the implementation of MRSDL described below modifies step 1 of the algorithm by selecting not only the analog A_i for which the estimated structural difference is least, but also all other analogs $A_j \ldots A_k$ whose estimated structural differences are within .05 of those of A_i. The actual closest structural match to the probe among $A_i, A_j \ldots A_k$ is then determined in step 2.

[2] Weighting the relations by their relative abundance in analogs was not found to increase retrieval accuracy.

Data Set	MRSDL		
	% exact	% close	comparisons
fables (80)	51.0	81.3	9.0
plays (17-18)	73.1	94.2	5.3
EA+ (11)	76.2	78.6	3.1
EA− (10)	78.6	92.9	2.5

Table 1: The proportion of MRSDL retrievals that were identical to the best match as determined by exhaustive match, the proportion of retrievals that returned an analog whose degree of match was within 5% of the closest analog, and the average number of structural comparisons required in each of the data sets. "EA+" and "EA−" represent instances and near-miss noninstances of employment activities, respectively.

Data Set	Surface Similarity		
	$AV_{p \in Probes}$ $Min\text{-}exact_p$	$MAX_{p \in Probes}$ $Min\text{-}exact_p$	$MAX_{p \in Probes}$ $Min\text{-}dl_p$
fables	14.9	71.3	68.0
plays	4.3	12.5	8.5
EA+	3.3	9.0	9.0
EA−	3.1	6.0	6.0

Table 2: Min-exact$_p$ is the minimum number of candidates that must be retrieved by surface similiarity to insure that the analog closest to probe p is in the candidate set. Min-dl$_p$ is the smallest candidate set size guaranteed to contain an analog whose degree of match is at least as great as the degree of match of the analog returned by MRSDL.

$SS(A)$ were distinct, then difference links were installed between $Ex(A)$ and $SS(A)$, and between A and $Ex(A)$. This approach was chosen because the number of difference links required is linear in the number of analogs and because the approach is consistent with a plausible scenario for acquiring knowledge of structural relations among analogs. A distinct configuration of difference links was constructed for each set of partitions used as source analogs.

In each of the retrieval approaches (except surface semantic retrieval) structure matching was performed by the best-first algorithm described in (Branting, 1991) running in greedy mode. To isolate the task of finding the structurally most similar analog from the contribution of various mapping constraints and to expedite the trials, the algorithm was run with information concerning semantic similarity among relations and case elaboration rules removed. Degree of structural similarity was measured by the proportion of propositions in the source analog that have an image in the target under the mapping that maximizes structural congruence.

Table one sets forth the performance of MRSDL averaged across four trials. The first column sets forth the proportion of MRSDL retrievals for each data set that were identical to the best match as determined by exhaustive search.[3] There are often several analogs having an almost identical degree of match with a probe. The second column sets forth the proportion of MRSDL retrievals that returned an analog whose degree of match was within 5% of that of the closest analog found by exhaustive search. The third column contains the average number of structural comparisons required for each MRSDL retrieval.

The first two columns of table two contain information concerning *Min-exact$_p$*, the minimum number of candidates that must be retrieved by surface similarity to insure that the analog closest to probe p is in the candidate set.[4] The first column sets forth the average of *Min-exact$_p$* for all probes p in each data set. This represents the average number of candidates that would be necessary for two-stage retrieval if one somehow knew *Min-exact$_p$* for every probe p. The second column sets forth the maximum of *Min-exact$_p$* for all probes p. This represents the smallest candidate set size that would guarantee for all probes that the candidate set would contain the best analog. The last column represents the maximum of *Min-dl$_p$*,[5] the smallest candidate set size guaranteed for all probes p to contain an analog whose degree of match at least as great as the degree of match of the analog returned by MRSDL.

Table three sets forth the average retrieval time in seconds of user CPU time for exhaustive search, BFIM, and MRSDL.

Discussion

Table one shows that MRSDL performs reasonably well, although not infallibly. In the fable and employment activity noninstance data sets MRSDL was over 90% accurate in retrieving analogs that were within 5% of the optimal match. Table two illustrates the shortcomings of two-stage retrieval. Although the average value of *Min-exact$_p$* was comparable to the average number of structural comparisons performed by MRSDL, each data set con-

[3] A separate comparison with BFIM was unnecessary because BFIM always finds the same match as exhaustive search

[4] If $Ex(p)$ is the analog found by exhaustive search of a given analog set with probe p and $SS(p,n)$ is the set of n closest surface matches to p, then *Min-exact$_p$* $= \min\{n \mid Ex(p) \in SS(p,n)\}$.

[5] If $Dl(p)$ is the analog found by MRSDL, *Min-dl$_p$* $= \min\{n \mid SS(p,n)$ contains some analog that matches p at least as well as $Dl(p)\}$.

Data Set	Exh.	BFIM	MRSDL
fables	10.7	2.9	5.0
plays	5.8	4.7	6.3
EA+	4.1	2.6	1.7
EA−	2.9	1.4	1.1

Table 3: Average retrieval times (in seconds of user CPU time) for exhaustive search, BFIM, and MRSDL.

tains some probe p for which $\mathit{Min\text{-}exact}_p$ is at least half the size of the data set. Thus, on these data sets at least, no two-stage retrieval scheme can simultaneous insure correctness and search less than half of the analogs in memory. The last column of table two illustrates that the smallest initial candidate set size guaranteed to equal the accuracy of MRSDL is at least half the size of the analog library.

Table three illustrates that MRSDL is usually substantially faster than exhaustive matching. The surprising exception was in the plays data set where MRSDL was actually slower than exhaustive matching. BFIM was consistently faster than exhaustive matching, more than three times as fast in the fable data set. Surprisingly, BFIM was also faster than MRSDL in two of the data sets. Parallel competitive matching has been criticized on grounds of psychological implausibility (Gentner and Forbus, 1991), but these data suggest that this retrieval technique can be relatively efficient.

Conclusion

This paper has presented a model of the use of expert knowledge to improve the accuracy of analog retrieval. This model, match refinement by structural difference links (MRSDL), is based upon the assumption that expertise in domains requiring analogical reasoning consists in part of knowledge of the structural similarities and differences between some pairs of the source analogs. In an empirical evaluation on four data sets, MRSDL generally found the most similar or nearly most similar source analog. Achieving comparable accuracy on these data sets with a two-stage retrieval technique such as MAC/FAC would require exhaustive matching with more than half of the source analogs.

References

Branting, L. K. (1991). *Integrating Rules and Precedents for Classification and Explanation: Automating Legal Analysis*. PhD thesis, University of Texas at Austin.

Branting, L. K. and Porter, B. W. (1991). Rules and precedents as complementary warrants. In *Proceedings of Ninth National Conference on Artificial Intelligence*, Anaheim. AAAI Press/MIT Press.

Falkenhainer, B., Forbus, K., and Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence Journal*, 41(1).

Faries, J. M. and Reiser, B. J. (1990). Terrorists and spoiled children: Retrieval of analogies for political arguments. To appear in *Proceedings of the AAAI Spring Symposium on Case-Based Reasoning*, Palo Alto, California, March 27–29.

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

Gentner, D. (1989). Finding the needle: Accessing and reasoning from prior cases. In *Proceedings of the Second DARPA Case-Based Reasoning Workshop*. Morgan Kaufmann.

Gentner, D. and Forbus, K. (1991). MAC/FAC: A model of similiarity-based retrieval. In *Thirteenth Annual Conference of the Cognitive Science Society*, pages 504–509.

Holyoak, K. and Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13(3).

Novick, L. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology*, 14:510–520.

Ross, B. (1989). Some psychological results on case-based reasoning. In *Proceedings of the Second DARPA Case-Based Reasoning Workshop*. Morgan Kaufmann.

Thagard, P., Holyoak, K., Nelson, G., and Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. Technical Report CSL-Report 41, Princeton University.

Wharton, C., Holyoak, K., Downing, P., Lange, T., and Wickens, T. (1991). Retrieval competition in memory for analogies. In *Thirteenth Annual Conference of the Cognitive Science Society*, pages 528–533.

Winston, P. H. (1980). Learning and reasoning by analogy. *Communications of the ACM*, 23(12).