Forbus, K. D., Gentner, D., Everett, J. O., & Wu, M. (1997). Towards a computational model of evaluating and using analogical inferences. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society,* 229-234.

Towards a Computational Model of Evaluating and Using Analogical Inferences

Kenneth D. Forbus
Institute for the Learning Sciences
John O. Everett
Institute for the Learning Sciences

Dedre Gentner
Psychology Department
Melissa Wu
Psychology Department

Northwestern University,
1890 Maple Avenue, Evanston, IL, 60201, USA

Abstract

Reasoning by analogy is a central phenomena in cognition. Existing computational models of analogy provide accounts of how analogical inferences are generated, but do not specify how they might be evaluated or integrated with other methods of reasoning. This paper extends the model of analogical inference in structure-mapping theory in two ways. First, we propose techniques for the structural evaluation of analogical inferences, to model one of the factors people appear to use in evaluating the plausibility of arguments based on comparisons. Second, we propose an information-level model of analogical inferences that supports reasoning about correspondences and mappings. We describe how this model fits with existing psychological evidence and illustrate its operation on several examples, using a computer simulation. These examples include evaluating the validity of a qualitative mental model and a prototype case-based coach that is being added to an already-fielded intelligent learning environment.

Introduction

Psychological results on analogical reasoning suggest that there are core techniques of comparison and analogical inference which, in concert with other processes, are used in tasks ranging from perception to conceptual change [Gentner & Markman 1997]. Yet there has been surprisingly little research on analogical inference. Existing computational models of analogy model the generation of *candidate inferences* (Gentner, 1982, 1983; Holyoak, Novick, & Melz, 1994; Keane, 1990), and particular simulations have used candidate inferences in modeling analogical reasoning and learning in physical domains (Falkenhainer, 1987; Forbus, Ferguson, & Gentner, 1994), but no models of analogical inference with a level of generality on par with models of mapping and retrieval have been proposed. Within the case-based reasoning (CBR) community analogical inference (or *adaptation*) has been one of the least explored aspects. Many effective CBR systems only act as retrieval systems, counting on human partners to understand and apply the retrieved information (Kolodner 1994; Schank & Cleary 1994). Those which do adaptation (c.f. Carbonell *et al* 1991; Kass 1986; Leake 1996) rely on domain-specific and task-specific methods. Our goal here is a domain-general account of how analogical inferences are evaluated and integrated with other knowledge.

Analogical inference is a complex phenomena, involving interactions of analogical processing with a variety of other cognitive processes. This paper provides another step towards a computational model of analogical inference, by extending Gentner's (1983) structure-mapping theory in two ways. First, we describe a method for *structural evaluation* of analogical inferences, which estimates how promising an inference is based on its form and the mapping that generated it. There is evidence that people use such estimates in deciding whether to pursue an analogy and which inferences are worth exploring further. Second, we describe a logical form for expressing analogical inferences that enables them to be integrated with other reasoning processes.

Our model is intended as a cognitive model in two respects: (1) The structural evaluation method should give answers that produce the same ordinal preferences in tasks as human subjects. (2) The logic of candidate inferences stands as an information-level model of the justifications someone would give for an analogical inference. We describe how this model is consistent with existing psychological evidence on analogical inference. We also illustrate how a computer simulation of the model can combine qualitative reasoning with analogical inferences to evaluate a possible analog for a home heating system, and how these techniques are being used in a prototype case-based coach for an intelligent learning environment.

Review of Structure-Mapping

According to structure-mapping theory, an analogy match takes as input two structured representations (*base* and *target*) and produces as output a set of mappings. Each mapping consists of a set of *correspondences* that align items in the base with items in the target and a set of *candidate inferences,* which are surmises about the target made on the basis of the base representation plus the correspondences. The constraints on the correspondences include *structural consistency,* i.e., that each item in the base maps to at most one item in the target and vice-versa (the *1:1 constraint*) and that if a correspondence between two statements is included in a mapping, then so must correspondences between its arguments (the *parallel connectivity* constraint). Which mapping is chosen is governed by the *systematicity* constraint: Preference is given to mappings that match systems of relations in the base and target. Each of these theoretical constraints is motivated by the role analogy plays in cognitive processing. The 1:1 and parallel connectivity constraints ensure that the candidate inferences of a mapping are well-defined. The systematicity constraint reflects a (tacit) preference for inferential power in analogical arguments.

The Structure-Mapping Engine (SME) (Falkenhainer *et al* 1986, 1989; Forbus *et al* 1994) is a cognitive simulation of analogical matching. Given base and target descriptions, SME finds globally consistent interpretations via a local-to-global match process. SME begins by proposing correspondences, called *match hypotheses,* in parallel between statements in the base and target. Then, SME filters out structurally inconsistent match hypotheses. Mutually consistent collections of match hypotheses are gathered into global mappings using a greedy merge algorithm. An evaluation procedure based on the systematicity principle is used to compute the *structural evaluation* for each match hypothesis and mapping. These numerical estimates are used both to guide the merge process and as one component in the evaluation of an analogy.

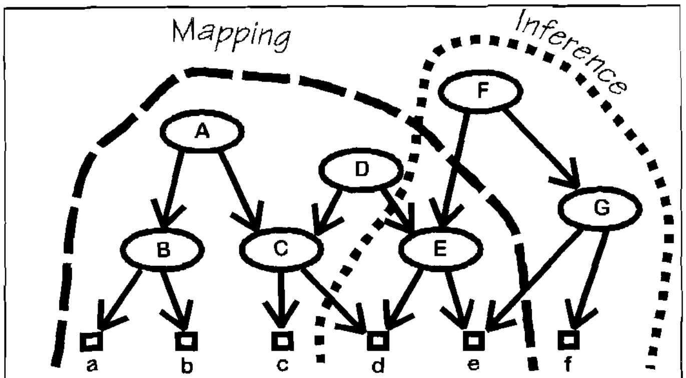

Figure 1: If propositions A-E and entities a-e are part of a mapping, then {F,G,E,e,f} would be used to generate a candidate inference. {E,e} are the "border" of the inference.

Candidate inferences for a mapping are generated by examining how the base intersects the mapping. Let a *root* of a description be a statement that is not the argument of any other statement in the description. If a root participates in the mapping, then it is part of the overlap between base and target, and can provide no new information. If a root is not part of the mapping, but has subexpressions that are part of the mapping, then a candidate inference is computed. In Figure 1, A, D, and F are roots. Since A and D participate in the mapping, only F serves as a starting point for a candidate inference. The form of the inference is the root expression, with substitutions made as necessary from the correspondences, and with skolem functions introduced for base constants that do not have correspondences (i.e., **f** in Figure 1).

Evaluating and Using Analogical Inferences

Structure-mapping theory defines analogical inferences as projections from the base that are structurally supported by the correspondences of a mapping. SME provides algorithms for automatically generating them. However, this is not enough: analogical inferences must be evaluated and used. Two central issues are:

1. How can the plausibility of analogical inferences be estimated? In addition to domain and task constraints, structural properties of the match are used by people as a factor in estimating plausibility.
2. What is the logical/explanatory import of analogical inferences? To capture their use in human reasoning, we must be able to couple analogical inference with other forms of reasoning.

We address each issue in turn.

Structural evaluation of candidate inferences

The structural evaluation of a mapping provides an estimate of match quality, based on the nature of the overlap. We suggest that a similar structural evaluation occurs psychologically for candidate inferences. However, for candidate inferences we postulate two distinct dimensions:

- *Support*: How much structural support does an analogical inference derive from the mapping that generated it?
- *Extrapolation:* How far does an analogical inference go beyond the support lent by the mapping?

We believe these two measures have significantly different functional roles. Support is the measure most like the structural evaluation of mappings: More is always better. Extrapolation is more complex: High extrapolation seems desirable in tasks like brainstorming or theory generation, but low extrapolation may be preferable for within-domain comparisons involving highly familiar situations. Consequently, we define each independently, although they are computed in a similar fashion.

The support and extrapolation structural evaluation algorithms are variations of the algorithm used for mappings. The score of a mapping is the sum of the scores of its correspondences. The scores of the correspondences are computed by the following algorithm: (1) each correspondence is given some initial score $\mathtt{Wi}$ and then (2) scores are incremented via a trickle-down method to enforce the systematicity preference for deep matching structures. That is, if $\mathtt{W(MH_1)}$ is the score associated with a match hypothesis $\mathtt{MH_1}$, $\mathtt{MH_2}$ is a match hypothesis that applies to one of MH_1's arguments, and δ is the trickle-down factor, then $\mathtt{W(MH_2)}$ is incremented as follows:

$$\mathtt{W(MH_2)} \leftarrow \mathtt{max\{W(MH_2)} + \delta\mathtt{W(MH_1);\ 1.0\}}$$

To compute the support score of a candidate inference, this same algorithm is used on the correspondences that support it in the mapping and adding up the results. Returning to Figure 1, the inference relies on the correspondences for E, d, and e, so the support score would be the sum of the scores for their correspondences. Notice that the support score encodes the systematicity preference: Trickle-down from E affects the scores for d and e.

The extrapolation score of an analogical inference is, roughly, the size of the new information over the total size of the inference. Consider two limiting cases. If there were no support (i.e., a hallucination), all the information would be new, so the extrapolation score would be 1. If there were nothing new (everything was there already), then the score would be 0. Any real candidate inference will be somewhere in between these two values.

The algorithm for computing extrapolation scores is

1. Apply the trickle-down algorithm used to score correspondences to the structure of the inference itself, i.e., as if we were matching the inference to itself
2. The extrapolation score is

$$\frac{\sum W(\text{outside})}{\sum W(\text{outside}) + \sum W(\text{inside})}$$

where inside refers to the items in the candidate inference that are part of the mapping and outside refers to the items in the candidate inference that are being projected. Again referring to Figure 1, the extrapolation score in this case would be

$$\frac{W(F) + W(G) + W(f)}{W(E) + W(d) + W(e) + W(F) + W(G) + W(f)}$$

Using the trickle-down algorithm provides a more conservative score than simply counting items would, since the existence of large structures outside the mapping will lead to higher scores inside the mapping due to trickle-down, although this effect is limited by the non-linearity of the trickle-down scheme, as noted above.

A logic of candidate inferences

The second requirement for evaluating candidate inferences is the ability to express them in a form that can interact with other processes. Let C be a candidate inference. We require C to be a proposition. Intuitively, the validity of C depends in part upon the validity of the correspondences that support it. We reify correspondences as propositions as follows. Let $\mathrm{MH}(b_1, t_1)$ be the hypothesis that b1 in the base corresponds to, i.e., matches, t1 in the target. The semantics of MH statements reflect the consistency constraints on match hypotheses, e.g., $\mathrm{MH}(b_1, t_1)$ is inconsistent with $\mathrm{MH}(b_1, t_2)$ and $\mathrm{MH}(b_2, t_1)$, for $b_1 \neq b_2$ and $t_1 \neq t_2$.

Analogical inference is not deductively valid[1]. That is, we may assume an analogical inference to be true in the absence of evidence to the contrary, but stand ready to retract it if it is implicated in a contradiction. Also, invalidating one analogical inference does not necessarily rule out other inferences made by the same mapping[2]. Therefore we must be able to express belief in the plausibility of each candidate inference independently. Let PLAUSIBLE-CI(C) be the proposition that candidate inference C is plausible, given available knowledge. PLAUSIBLE-CI is nonmonotonic, in the same sense of McCarthy's (1987) ABNORMAL predicate or Hobbs *et al* (1993) ETC predicate. That is, we assume that in reasoning PLAUSIBLE-CI statements are assumed to be true in the absence of information to the contrary, but will be viewed as candidates for retraction if contradictions arise.

Given these definitions, we can now express an analogical inference as follows:

$$\mathrm{MH}(b_1, t_1) \& \ldots \& \mathrm{MH}(b_n, t_n) \& \mathrm{PLAUSIBLE\text{-}CI}(C) \Rightarrow C$$

That is, the candidate inference is justified by the correspondences between the base and target, unless it is discovered to be invalid.

A further piece of vocabulary is needed in order to capture the intuition that belief in a candidate inference is tied to belief in the mapping that generated it. The importance of structural consistency in analogical reasoning suggests that people work with mappings rather than isolated correspondences. The predicate USING-MAPPING serves as a control assertion indicating belief in the set of correspondences structurally entailed by a mapping. Each USING-MAPPING statement justifies MH statements concerning its correspondences, i.e., if mapping M pairs items $b_1, t_1 \ldots b_p, t_p$, then

$$\mathrm{USING\text{-}MAPPING}(M) \Rightarrow \mathrm{MH}(b_1, t_1)$$

...

$$\mathrm{USING\text{-}MAPPING}(M) \Rightarrow \mathrm{MH}(b_p, t_p)$$

With this vocabulary the results of analogical matches can be expressed in a form that captures our intuitions about the structural dependencies of a candidate inference. For purposes of simulation, this vocabulary can be used to express the results of analogical processing in a form that can be used by other processes. This facilitates the simulation of tasks that use analogical inferences in combination with other reasoning processes.

Implementation

We have extended SME to compute the support and extrapolation scores for candidate inferences. We integrated SME with the LTRE reasoning system from Forbus & de Kleer (1993). Promising analogical inferences (based on task-specific criteria) are installed in the LTRE according to the logic described previously, and assuming the PLAUSIBLE-CI statement to be true by default.

Psychological Support for the Model

The ability to identify which inferences follow from a given set of correspondences has been demonstrated experimentally (Clement & Gentner 1991; Spellman & Holyoak 1996). Markman (in preparation) has demonstrated that analogical inferences follow structural consistency, even when there are multiple possible mappings. Our model for the logical form of analogical inference is consistent with these results.

Our definition of support score is consistent with several lines of evidence. Psychologically, matches involving larger systems of statements are viewed by subjects as more sound (Gentner *et al* 1993). Clement & Gentner (1991) showed that subjects made predictions based on statements connected to a common antecedent in the base, and that candidate inferences connected to systematic base structures are preferred to those which are not.

Similarity has been suggested as a central process in induction tasks (Heit & Rubenstein, 1994; Lassaline 1996; Osherson *et al* 1990), so it is useful to see how this model fits with these studies. Lassaline (1996) asked subjects to rate the similarity of pairs of fictitious animals and the inductive strength of a property inference (i.e., if A has X, W, and Z, and B has X, Y, and Z, how likely is it that A has Y?, where A and B were fictitious animals and the rest of the variables were filled in with properties such as "dry flaky skin" or "attacks of paranoia") . Adding a relation in the base that explained the inferred property (i.e., telling the subjects that in B, X causes Y while leaving the description of A unchanged) increased inductive strength, but adding a relation that was not connected to the inference did not. A simple model of this task is to treat it as analogical mapping, with animal B serving as base and animal A as the target,

[1] See Falkenhainer (1990) for a discussion of the relationship between analogy, deduction, abduction, and induction.

[2] One prediction of the atom/solar system analogy is that electrons would have "moons" orbiting them. The failure to find these moons did not cause the analogy to be abandoned.

and treating inductive strength as a function of the candidate inference support score. Using these assumptions, a simulation of her experiments also yields this result[3]. Figure 2 illustrates.

```
=========================
Base description ex2-la-lr-B
(CAUSE (DRY-FLAKY-SKIN B) (FREQUENT-HEADACHES B))
(RESTLESS-SLEEPING-HABITS B)
(ACTIVE-METABOLISM B)
=========================
Target description ex2-la-lr-A
(RESTLESS-SLEEPING-HABITS A)
(ACTIVE-METABOLISM A)
(DRY-FLAKY-SKIN A)
========================
Mappings for <SME 4>.
   ;; Mapping 5: Score  0.4950
[Correspondences omitted]
Inferences:
(CAUSE (DRY-FLAKY-SKIN A) (FREQUENT-HEADACHES A))
Support = 1.082
Extrapolation = 0.520602569782898
```

Figure 2: An example of simulating inductive inference via analogical mapping.

Heit and Rubenstein (1994) found that people make stronger inferences about whether one animal has a property based on another animal's having it when the kind of property to be inferred (anatomical or behavioral) matches the kind of similarity between the animals (anatomical or behavioral). For instance, people judge the likelihood that whales travel shorter distances in extreme heat to be higher when told that tuna do, relative to when they are told that bears do, presumably because whales and tuna have more behavioral overlap (both swim) while whales and bears match anatomically (both mammals). If subjects are linking new properties with their existing knowledge about these animals, then this result is consistent with our model because the set of support for the analogical inference would be higher.

Osherson *et al* (1990) investigated multi-premise inductive arguments, i.e., robins use serotonin as a neurotransmitter, bluejays use serotonin as a neurotransmitter, therefore sparrows use serotonin as a neurotransmitter, modeling them as category-based induction. They give similarity a central role in their model, but assume only that a numerical value for the similarity of two objects (again, animals) can be computed. The structural evaluation of a mapping could serve this purpose. If we further assume a SEQL-like model of abstraction from multiple comparisons (Skorstad, Gentner, & Medin, 1988), analogical inference may also play a larger role in explaining some of the same phenomena. For example, Osherson *et al* found that an argument that a property held for a category was stronger if the premises involved more typical members of a category. The structural descriptions for more typical members of a category might have more overlap among themselves than descriptions of a set of less-typical members, thus they might provide more support for candidate inferences.

```
Inference:
(implies
 (and (continuous-settable-control thermostat)
      (controls (setting thermostat) furnace))
 (qprop (applied-heat furnace)
        (setting thermostat)))
Support = 6.0
Extrapolation = 0.273607748184019
[...details omitted...]
Contradiction found for
 #<CANDIDATE-INFERENCE #x10799D8>
  of <Mapping 32>:
1. Setting of THERMOSTAT controls FURNACE.
2. THERMOSTAT is a continuously settable control.
3. applied heat of FURNACE has no indirect
   influences.
Retracting #<CANDIDATE-INFERENCE #x10799D8>.
```

Figure 3: Mental model denied.

Examples

We have tested our model of analogical inference on a variety of examples. We describe two examples next to illustrate that it can operate on complex representations, including automatically generated descriptions.

Evaluation of potential analogs

A common misconception about home heating systems is that, if your house is cold, setting your thermostat to a higher setting than ultimately desired will cause it warm up faster. Kempton (1986) showed in interviews that this faulty mental model is often due to mistaken analogies, such as a gas pedal analogy. Pushing the pedal down farther causes the car to reach the desired speed sooner because the engine will supply more power to the wheels. Adopting this analogy typically leads to higher heating bills without increased comfort, since the temperature will not increase any faster, and eventually will overshoot and must be turned down.

How might someone escape from this mistaken analogy? The ability to integrate analogical inference with other forms of reasoning enables us to model the process of evaluating such analogies. It is well-known that in common sense reasoning it is virtually impossible to have a complete set of antecedents for conclusions (McCarthy 1987). Thus in the gas pedal scenario, the control relationship between the engine and the continuous nature of the gas pedal's setting might be conjectured to be sufficient to provide continuous control over the engine, which means that more throttle leads to faster attainment of a desired speed. Using SME to compare two representative descriptions, a single mapping is generated whose candidate inference is shown in Figure 3. The home heating scenario, with this inference, is then analyzed using a qualitative physics system (Forbus, 1984) on the same description database that SME used. After instantiating a simple domain theory, finding what physical processes were acting, and resolving influences to figure out the

[3] Lassaline also found that adding shared attributes increased inductive strength, while adding shared relations did not. Our model does not exhibit this behavior. However, other experiments have found, consistent with our model, that inductive strength increases with similarity (Osherson *et al* 1990).

causal dependencies between quantities, a contradiction was found because, according to the domain theory, the applied heat of the furnace is an independent parameter. (This explanation is also shown in Figure 3)

Case-based Coaching in Education

CBR systems are sometimes used in educational software as a coach, to support students doing a task (e.g., Schank & Cleary 1994). We are adding a software design coach to CyclePad (Forbus & Whalley, 1994), an intelligent learning environment for engineering thermodynamics. CyclePad is based on the idea of teaching principles by engaging students in design tasks, such as designing power plants, aircraft engines, and refrigerators. CyclePad is currently in experimental use by students at Northwestern University, Oxford, and the US Naval Academy. A recurring problem is that students, being novice designers, tend to get stuck. If their design fails to meet its requirements, how might they improve it? Case-based coaching is a natural approach for this task, since basing advice on interesting examples gives students additional motivation and context.

We have used our analogical inference system to create a prototype case-based coach module for CyclePad. The idea is this: When the student asks for help, the current state of their design is augmented with a teleological description generated by CARNOT (Everett, 1995), a program that recognizes the intended purpose of the parts of the cycle and how they are relevant to the student's goals. This description is used to retrieve a case from a library of designs that solves a similar problem. Using analogical inference to adapt this example to the student's particular problem, the coach will then offer concrete advice on how the student might improve their design, using the case as its justification. (See Fig. 4) With the exception of automatic linking of task-specific criteria and connecting it to the existing CyclePad interface, this coach has been fully implemented.

The case library is directed at design problems students are likely to have. Entries in the case library are created as follows: The domain expert uses CyclePad to construct a cycle that illustrates a particular problem. In "watch me" mode, the expert then modifies the design in a way that fixes the problem. Thus the structural description of the cycle, CARNOT's teleological analysis of what the cycle does and how each part of the cycle contributes to this function, and a

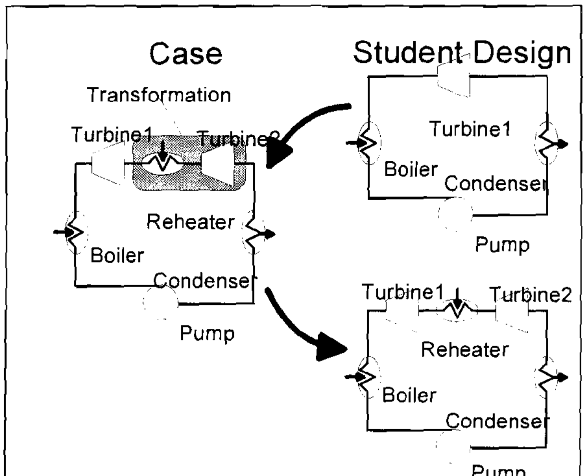

Figure 4: Case-based coaching provides design hints

formal representation of the expert's transformation are all automatically generated for the case. The only hand-input part of the representation is the expert's specification of the exact nature of the problem, i.e., low thermal efficiency or high operating cost, which has to be stated in a tightly constrained formal representation language and added to the case. The selection of 12 initial cases was based on the likely needs of intermediate thermodynamics students. The test problems were generated in the same way. The average number of expressions in each case is 77 and the average number of entities is 19.

We use the MAC/FAC retrieval model [Forbus *et al* 1995] to retrieve cases. MAC/FAC output's is further filtered as follows: Any candidate inference that does not hypothesize a transformation is eliminated from consideration as irrelevant. (There can be at most one relevant inference per retrieved case, but sometimes there are one or two extra, irrelevant inferences, and sometimes a retrieved case does not yield a useful inference.) For those remaining, the inference with the highest support score is chosen as the advice to give to the student.

In our experiments so far, we have found that when multiple cases were retrieved, choosing the analogical inference with the highest support score always provides the optimal advice. This result should be viewed with caution, since the number of problems tried has been small, the case base is only about one-fourth of what we believe is needed for broad coverage, and, most importantly, it has not been field-tested with students. Even so, this does suggest that general-purpose cognitive simulation tools, operating on rich, automatically generated case libraries, can provide accurate and efficient case-based coaching.

Related Work

In addition to the case-based reasoning work mentioned earlier, there are a variety of cognitive models of analogical mapping and retrieval. Existing models of general analogical processing (such as Keane's (1990) IAM, Holyoak & Thagard's (1989) ACME, and Hummel & Holyoak's (in press) LISA) often provide methods for generating inferences. However, they do not provide evaluation methods or integrate them with other reasoning systems. Furthermore, ACME and LISA do not guarantee to produce structurally consistent mappings, which makes it difficult to get accurate analogical inferences (Gentner, 1982; Markman, in preparation). Falkenhainer's (1990) PHINEAS used SME in a model of scientific discovery. Its techniques for using and evaluating analogical inferences were specific to its task and domain.

Discussion & Future Work

Analogical inference is a complex phenomenon to model because it involves the interaction of a number of cognitive processes. In this paper we extended the structure-mapping notion of candidate inferences in two ways. First, we proposed a method for the structural evaluation of candidate inferences. This allows evaluating the goodness of the inference in terms of its relation to the mapping that generated it. Second, we developed a vocabulary for logically expressing the relationship between a candidate inference and the structural correspondences that support it. These extensions appear compatible with the overall pattern of psychological results on analogical inference. Moreover, we have demonstrated that they can be implemented effectively and used in systems that combine analogical reasoning with other forms of reasoning and can operate with complex, automatically generated representations[4].

We are exploring this model further in two ways: (1) we are designing experiments to test the psychological plausibility of the extrapolation score and other predictions of the model and (2) we are integrating this model into a new cognitive simulation of analogical problem solving and reasoning. We hope that these extensions bring us a step closer towards a full computational model of analogical inference.

Acknowledgments

We thank Art Markman, Ron Ferguson, Jeff Lowenstein, and the Similarity and Analogy group at Northwestern for valuable discussions. This research was supported by the Cognitive Science Division of the Office of Naval Research, Contract N00014-92-J-1098.

References

Carbonell, J., Knoblock, C. and Minton, S. 1991 PRODIGY: An integrated architecture for planning and learning. In *Architectures for Intelligence*, VanLehn, K. (Ed.). Erlbaum.

Clement, C. A. & Gentner, D. 1991. Systematicity as a selection constraint in analogical mapping. Cognitive Science, 15, 89-132.

Everett, J. O. 1995. A Theory of Mapping from Structure to Function Applied to Engineering Domains. 14th International Joint Conference on Artificial Intelligence, Montreal, Morgan Kaufmann.

Falkenhainer, B., Forbus, K., and Gentner, D. 1986. The Structure-Mapping Engine" Proceedings of AAAI-86, Philadelphia, PA, August.

Falkenhainer, B., Forbus, K. D., & Gentner, D. 1989. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, **41**(1), 1-63.

Falkenhainer, B. 1990. A unified approach to explanation and theory formation. In Shrager & Langley, editors, Computational Models of Scientific Discovery and Theory Formation. San Mateo, CA: Morgan Kaufmann.

Forbus, K. D. 1984. Qualitative process theory. Artificial Intelligence, 24(1), 85-168.

Forbus, K. and de Kleer, J. 1993. *Building Problem Solvers*, MIT Press.

Forbus, K. D., Ferguson, R. W., and Gentner, D. 1994. Incremental Structure Mapping. In Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Forbus, K. D., Gentner, D., & Law B.K. 1995. MAC/FAC: A model of similarity-based retrieval. Cognitive Science, 19(2), 141-205.

Forbus, K. and Whalley, P. 1994. Using qualitative physics to build articulate software for thermodynamics education. *Proceedings of AAAI-94*, Seattle

Gentner, D. 1982. Are scientific analogies metaphors?. In Miall, D. (Ed.) *Metaphor: Problems and perspectives.* Harvester Press Ltd.

Gentner, D. 1983. Structure-mapping: a theoretical framework for analogy. Cognitive Science, 23, 155-170.

Gentner, D., & Markman, A. B. 1997. Structural alignment in analogy and similarity. American Psychologist

Gentner, D., Rattermann, M.J., & Forbus, K.D. 1993. The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology* 25, 524-575.

Hobbs, J., Stickel, M., Appelt, D. and Martin, P. 1993. Interpretation as Abduction. *Artificial Intelligence.* 63:1-2, 69-142.

Holyoak, K. J., & Thagard, P. 1989. Analogical mapping by constraint satisfaction. Cognitive Science, 13(3), 295-355.

Holyoak, K.J. & Thagard, P. 1995. Mental leaps: Analogy in creative thought. Cambridge, MA: The MIT Press.

Hummel, J.E., & Holyoak, K.J. (in press). Psychological Review.

Kass, A. 1986. Modifying explanations to understand stories. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Erlbaum.

Keane, M. T. G. (1990). Incremental analogizing: Theory and model. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), Lines of Thinking London: John Wiley and Sons, Ltd.

Kempton, W. 1986. Two theories of home heat control. *Cognitive Science* 10:75-90.

Kolodner, J. L. (1994). Case-based reasoning. San Mateo, CA: Morgan Kaufmann Publishers.

Leake, D. (Ed.) 1996. *Case-Based Reasoning: Experiences, Lessons, and Future Directions*, MIT Press.

Markman, A.B. in preparation Constraints on analogical inference.

Medin, D. L., Goldstone, R. L., & Gentner, D. 1993. Respects for similarity. Psychological Review, 100(2), 254-278.

McCarthy, J. 1987. Circumscription: A form of nonmonotonic reasoning., in Ginsberg, M. (Ed.) *Readings in Nonmonotonic Reasoning*, pp 145-152, Morgan Kaufmann.

Schank, R. and Cleary, C. 1994. *Engines for Education.* Erlbaum.

Skorstad, J., Gentner, D., & Medin, D. (1988). Abstraction processes during concept learning: A structural view. In Proceedings of the Tenth Annual Conference of the Cognitive Science Society. Montreal: Lawrence Erlbaum Associates.

---

4 We are incorporating the case-based coach for CyclePad described here into the next major release of the system for field-testing with students.