Cognitive Modeling of Analogy Events in Physics Problem Solving From Examples

Matthew Klenk (m-klenk@northwestern.edu)
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL 60208 USA

Ken Forbus (forbus@northwestern.edu)
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL 60208 USA

Abstract

Understanding how analogy is used in problem solving is an important problem in cognitive science. This paper describes a model of using worked solutions to solve new problems, in terms of structure-mapping processes in the Companions cognitive architecture. The Educational Testing Service independently evaluated the flexibility of the system by using AP Physics problems that were systematically varied to test different types of transfer. We also show that the model provides an explanation for many of the analogy events in VanLehn's (1998) analysis of the use of analogy by students solving physics problems.

Keywords: Analogy; Problem Solving

Introduction

Cognitive science research has shown that analogy plays important roles in problem solving and learning (Gentner & Gentner, 1983; Holyoak 1985; Ross 1987; Novick 1988). One role is facilitating problem solving by using worked solutions. For example, VanLehn and Jones (1993a) observed that students used analogical reasoning in solving physics problems even when the underlying first principles knowledge was already known and had been successfully used before. This paper describes how the Companions cognitive architecture (Forbus & Hinrichs 2006) models example use in solving AP Physics problems. The AP Physics exam is taken by high school students in order to receive college credit. This is an interesting domain because students find such problems difficult and there is a wealth of prior cognitive science solving research on physics problem. Figure 1 shows four examples generated for this work by the Educational Testing Service (ETS), the company which administers the AP Physics exam.

We start by briefly reviewing the Companions architecture, focusing on the key analogical processes used. Then, we describe how a Companion uses these processes to solve physics problems. Next, the results of an external evaluation demonstrating the system's ability to successfully transfer across different problem variations are summarized. This is followed by an analysis of the 90 problems from this evaluation in terms of analogy events (as defined by VanLehn (1998)), showing a qualitative match with the patterns of analogy events found in protocols. Finally, we discuss other related work and future plans.

The Companions Architecture

The Companions architecture is exploring the hypothesis that structure-mapping operations (Gentner 1983; Forbus & Gentner 1997) are important building blocks for modeling reasoning and learning. This hypothesis suggests that within domain analogies, where new situations are understood in terms of prior understood examples, provide an important source of breadth and robustness of human common sense reasoning. Forbus & Hinrichs (2006) provides an overview of the Companions architecture; for this paper, the key processes to understand are analogical matching and retrieval. We summarize each in turn.

The Structure-Mapping Engine (SME) models the structure-mapping process of comparison (Falkenhainer, Forbus, & Gentner 1989). Structure-mapping postulates that this process operates over two structured representations (the *base* and *target*), and produces one or more *mappings*,

1. A ball is released from rest from the top of a 200 m tall building on Earth and falls to the ground. If air resistance is negligible, which of the following is most nearly equal to the distance the ball falls during the first 4 s after it is released? (a) 20m; (b) 40m; (c) 80m; (d) 160m.
2. An astronaut on a planet with no atmosphere throws a ball upward from near ground level with an initial speed of 4.0 m/s. If the ball rises to a maximum height of 5.0 m, what is the acceleration due to gravity on this planet? (a) 0.8m/s^2; (b) 1.2m/s^2; (c) 1.6m/s^2; (d) 20m/s^2;
3. A box of mass 8kg is at rest on the floor when it is pulled vertically upward by a cord attached to the object. If the tension in the cord is 104N, which of the following describes the motion, if any, of the box? (a) It does not move; (b) It moves upward with constant velocity; (c) It moves upward with increasing velocity but constant acceleration; (d) It moves upward with increasing velocity and increasing acceleration.
4. A block of mass M is released from rest at the top of an inclined plane, which has length L and makes an angle q with the horizontal. Although there is friction between the block and the plane, the block slides with increasing speed. If the block has speed v when it reaches the bottom of the plane, what is the magnitude of the frictional force on the block as it slides? (a) $f = Mg\sin(q)$; (b) $f = Mg\cos(q)$; (c) $f = MgL\sin(q) - ½Mv^2$;(d) $f = [MgL\sin(q) - ½Mv^2]/2$.

Figure 1: Example AP Physics problems

Klenk, M. and Forbus, K. 2007. Cognitive modeling of analogy events in physics problem solving from examples. In *The Proceedings of CogSci-07*. Nashville, TN.

each representing a construal of what items (entities, expressions) in the base go with what items in the target. This construal is represented by a set of *correspondences*. Mappings also include a *score* indicating the strength of the match, and *candidate inferences* which are expressions from the base which, while unmapped in their entirety, have subcomponents that participate in the mapping's correspondences. SME operates in polynomial time, using a greedy algorithm (Forbus & Oblinger 1990).

MAC/FAC (Forbus *et al.* 1994) models similarity-based retrieval given a case of facts, or *probe*, and a large case library. The first stage, using a special kind of feature vector automatically computed from structural descriptions, rapidly selects a few (typically three) candidates from the case library. The second stage uses SME to compare these candidates to the probe, returning one candidate reminding (or more, if they are very close) for the probe.

As performance systems, both SME and MAC/FAC have been used successfully in a variety of different domains, and as cognitive models, both have been used to account for a variety of psychological results (Forbus 2001).

Solving Problems by Worked Solutions

As noted above, students commonly use analogies with worked solutions to solve physics problems. Our Companion's model starts with some basic mathematical skills, a broad common sense ontology and some qualitative mechanics, representing the background knowledge that a student might bring to such problems. The representations use the ontology of the ResearchCyc knowledge base, plus our own extensions. ResearchCyc is useful because it includes over 30,000 distinct types of entities, over 8,000 relationships and functions, and 1.2 million facts constraining them. Thus, everyday concepts like "throwing" and "ball" are already defined, rather than us generating them specifically for the purpose of this project.

Unlike many students, our model does not start with any knowledge of the equations of physics. All equations and knowledge of when to apply them come from analogies with worked solutions, worked through example problems found in textbooks. While not representative of many students, it provides an interesting extreme assumption for measuring the power of analogy in its purest form.

The problems and worked solutions used throughout this work were generated by the Educational Testing Service, the company which administers the AP Physics exam. The representation conventions used were established in collaboration with us and with Cycorp, the creators of the ResearchCyc KB contents. ETS then generated all problems and worked solutions using templates, which were not available to us.

```
 (groundOf Planet-1 Ground-1)
(performedBy Throwing-1 Astronaut-1)
(no-GenQuantRelnFrom
        in-ImmersedFully Planet-1 Atmosphere)
(eventOccursNear Throwing-1 Ground-1)
(objectThrown Throwing-1 Ball-1)
(querySentenceOfQuery Query-1
        (valueOf (AccGravityFn Planet-1) Acc-1)
…
```

Figure 2: Part of Problem 2 representation

Example Problem and Worked Solution

Figure 2 shows part of the 37 facts used to represent Problem 2 from Figure 1. The worked solutions were created at roughly the level found in textbooks. They are not deductive proofs, nor problem solving traces in the language of our solver. This is important, because it provides more opportunities for the system to learn (and to make mistakes). For Problem 2 from Figure 1, the worked solution consisted of 7 steps.

1. Categorize the problem as distance-velocity problem under constant acceleration
2. Instantiate the distance-velocity equation ($V_f^2 = V_i^2 - 2ad$)
3. Given the projectile motion and lack of atmosphere, infer that the acceleration is equal to the acceleration due to gravity ($a = g$)
4. Because of the projectile motion event, the bat is not moving at the maximum height ($V_f = 0$ m/s)
5. Solve the equations for the acceleration due to gravity ($g = -1.6$ m/s^2)
6. Sanity check the answer (the answer is consistent)
7. Select the appropriate multiple choice answer ("c")

```
(stepType Step3 DeterminingValueFromContext)
(stepUses Step3 (isa Throwing-1 ThrowingAnObject))
(stepUses Step3 (occursNear Throwing-1 Ground-1))
(stepUses Step3
     (no-GenQuantRelnFrom
         in-ImmersedFully Planet-1 Atmosphere))
(stepUses Step3 (objectMoving Upward-1 Ball-1))
…
(stepUses Step3 (direction Upward-1 Up-Directly))
(solutionStepResult Step3
     (valueOf
       (AtFn ((QPQuantityFn Speed) Ball-1)
              (EndFn Upward-1))
       (MetersPerSecond 0)))
```

Figure 3: Problem 2 Worked Solution Step 3

Figure 3 illustrates how step 3 is represented. Worked solutions are stored along with the problem description as a case in the Companion's case library, so they are available when solving new problems.

Solving Problems via Analogy

Problems are presented as cases. The first phase of problem solving is to generate an analogy with a relevant example. This is done in three steps. First, the Companion uses MAC/FAC to try to retrieve a worked solution from the case library, using the problem as the probe. The mapping is examined to see whether or not it includes all of the problem's event structure, i.e., the physical events described

Klenk, M. and Forbus, K. 2007. Cognitive modeling of analogy events in physics problem solving from examples. In *The Proceedings of CogSci-07*. Nashville, TN.

in the problem, such as motion. If not all events are mapped, MAC/FAC is used again, but with only the unmapped events as the probe. This process continues until either there is no more unmapped event structure or no more matches are found. The result is one or more examples that should provide relevant facts about the new problem, including equations. Problem solving proceeds by mining the candidate inferences of these mappings.

The system begins by using rules to categorize the problem and determine which quantity or quantities should be solved for. (In Problem 3, for example, determining which answer is consistent requires finding values for velocity and acceleration.) Values for quantities are found in three ways. (1) The value might already be known as part of the problem. (2) The candidate inferences of the mapping may contain a solution step in which the goal parameter was assumed in the worked solution. (3) The candidate inferences provide an equation containing the sought quantity. In this case, the system first looks for values for the other quantities in the equations, and then attempts to solve the equation for the original parameter. The algebra routines are straightforward, based on the system in Forbus & de Kleer (1993). We currently treat the mathematical operations involved in solving a problem as a black box, not subject to learning.

To determine whether or not a solution step suggested by candidate inferences is valid, the system checks its context in the worked solution. Suppose for example the step assumes that the acceleration of a rock in free fall is 10 m/s^2, because the rock is falling on Earth and the problem says to ignore air resistance. To apply this step, the system must be able to infer that there is no air resistance in the current situation and that the event occurs on Earth. This verification helps guard against inappropriate applications of solution steps.

Before selecting a multiple choice answer, the system searches the candidate inferences for any solution steps representing sanity checks. For example, a problem asking, "How far would a ball fall off a 200m building in 4s?" would have a sanity checking step in which the answer, 80m, was compared to the height of the building, 200m. Since the answer is less than the height of the building, the result of the step is that the ball fell 80 meters. When this worked solution is used in solving future problems, the analogy produces candidate inferences indicating the type of check and corresponding quantities in the current problem that are involved. Currently, we only employ this check if the quantity sought for is involved in the comparison. This is because it is clear how to resolve a failure, i.e. use the value compared against it instead, because it constitutes a limit point (Forbus 1984) for that situation. This is a reasonable heuristic for Mechanics but in other situations and domains what to do is much less clear, and we plan to learn rules for resolving such problems in the future.

After an answer is found to be consistent, it is compared against each of the answer choices. The system selects either the closest answer for quantity value questions or the consistent answer choice in a qualitative behavior problem, such as Problem 3 in Figure 1.

Experimental Results

ETS conducted a formal evaluation of the system's problem solving performance, and we collected additional information on analogy events afterwards. ETS presented quizzes and worked solutions to a Companion running on a cluster remotely. The format of the questions mirrors the Chi *et al.* (1989) study by looking at problem solving over systematic variations of mechanics problems. The Chi *et al.* study involved giving the subjects worked solutions of three different problem types and then testing them on four isomorphic problems of each type followed by a second set of problems from the same chapter but not isomorphic. First, we describe the ETS evaluation of the system over systematic variations of problems, called here transfer levels. Then, we compare the distribution of the analogy events in this study to those reported by VanLehn's (1998) analysis of the Chi *et al.* protocols.

Problem Solving Across Transfer Levels

Our study involved having five worked examples of each of the four different problem types for a total of 20 problems in memory. Then the system was presented quizzes of systematic variations of problems representing different transfer levels as follows[1]:

1. *Parameterization*: changing the parameter values, but not qualitative outcome
2. *Extrapolation*: changing the parameters such that the qualitative outcome changes as well
3. *Restructuring*: asking for a different parameter
4. *Extending*: including distracting information
5. *Restyling*: changing the types of everyday objects involved
6. *Composing*: requiring concepts from multiple problems

The problems and their worked solutions were created by ETS in collaboration with Cycorp. The authors saw less than half of the transfer problems prior to the evaluation. Each transfer level (TL) contained 5 problem variations of each type making 20 total problems per TL. TL-2 only contained 10 problems because two of the problem types could not be altered to produce qualitatively different outcomes, Problems 2 and 4 from Figure 1. In each transfer level, every problem had a transformation based upon the transfer level to a specific worked example in memory. The data presented here is a representative subset of the whole evaluation, which also investigated learning rates.

Table 1 presents the system's performance by transfer level. The significance is computed against random chance on a 4 option multiple choice test. Given the fact that this evaluation was conducted externally on unseen problems

[1] These levels are from a 10-level catalog of transfer tasks used in DARPA's Transfer Learning Program (http://fs1.fbo.gov/EPSData/ODA/Synopses/4965/BAA05-29/BAA05-29TransferLearningPIP.doc)

Table 1: Accuracy Results

Transfer Level (# of problems)	Percentage Correct
1 – Parameterization (20)	90% (p<.05)
2 – Extrapolation (10)	50% (p<.10)
3 – Restructuring (20)	25% (p=.58)
4 – Extending (20)	90% (p<.05)
5 – Restyling (20)	95% (p<.05)
6 – Composing (20)	45% (p<.05)

and our analysis of its errors, the system performed quite well. Four out of the six transfer levels were statistically significant (p<.05). For TL-1, TL-4, and TL-5, the system scored at least 90% correct. Our rates were not as high for the other levels; TL-2 was 50%, TL-3 was 25%, and TL-6 was 44%. Even when the system failed to produce the correct answer, the retrieval algorithm always selected the correct problem or problems to reason from. This mirrors the findings of VanLehn and Jones (1993b) where human subjects rarely referenced examples that were not maximally similar to the problem.

While the results on three of the levels illustrate the strengths of the analogical approach, there are some results that require more explanation. First, in TL-2, the representations concerning the worked solution for problem type 3 incorrectly formulate the sanity checking step in which if a negative acceleration is calculated, the acceleration is inferred to be zero. We were unable to correct this given the external nature of the evaluation. The errors in TL-3 and TL-6, where the Companion was unable to score above 50 percent on any of the quizzes, were due to limitations in the problem solver's strategies. For Problem Type 3, the solver did not handle substituting different parameter values for answer choices efficiently enough to prevent timeouts. The low scores on TL-6 are because the solver's strategies assume that a given problem either demands numerical values or symbolic values, but not both, and thus it could not handle a composition of a symbolic problem with a numerical problem. Given the system's focus on transferring domain knowledge, it could not overcome these problems. Future work in learning problem solving strategies and interactivity is motivated by these results.

Analysis of Analogy Events

Chi *et al.* (1989) collected verbal protocols from 9 subjects learning Newtonian Mechanics while they studied worked examples and solved a series of problems. These protocols were used to investigate the differences between good and poor problem-solvers, including creating the Cascade system, which modeled the self-explanation effect (VanLehn *et al.* 1992). When attempting to fit Cascade to this data, VanLehn and Jones (1993) observed that people used analogical reasoning even they could have used first-principles knowledge. Later, VanLehn (1998) reanalyzed the original protocols, leading to a new taxonomy of analogy events:

- *Initialization events* – the subject sets up a mapping between the examples and the problems.
- *Transfer events* – the subject infers something about the solution from an example. These events were further divided by the type of inference made:
  1. *Line*: The subject transferred a whole equation, vector, or diagram.
  2. *Part of a line*: The subject transferred a detail from a line, such as whether a projection function was sine or cosine, or whether a vector went up or down.
  3. *Search control*: The subject made the decision on what steps to do by consulting the example and seeing what steps it did.
  4. *Checking*: The subject decided whether their most recent action or decision was correct by consulting the example.
  5. *Failure*: The subject failed to find anything useful during this transfer event.

Initialization events were indicated by the subject flipping the book to a worked solution, reading some of the example and deciding if it will be useful to solve the current problem. VanLehn found that initialization events usually occur at the beginning of the problem solving, consistent with (Bassok & Holyoak 1989; Faries & Reiser 1988; Ross 1989).

Most of these events can easily be mapped to our model. In our model, initialization events occur at the beginning of problem solving when the system uses MAC/FAC to retrieve a similar example(s) from memory. Our use of recursive retrievals for complex analogs is consistent with VanLehn's finding that for complex problems, there could be multiple initialization events.

Because our model answers all problems using analogy, it relies heavily on what would be construed as transfer events in VanLehn's analysis. Line transfers are indicated in our model in two ways: (1) Using a candidate inference to map an equation from the worked solution onto the problem and (2) inferring parameter values from the problem situation. Consistent with VanLehn's findings, these events occur throughout the problem solving process.

While similar in some ways, there are deep differences between our model and Cascade. In Cascade, line transfer events are only used when rule-based reasoning reaches an impasse. Cascade modeled analogical search control by storing triples containing the problem, goal, and the rule used to achieve the goal. The triples represent an approximation of episodic memory. When the system faced search control decisions in the future, it would look for a similar goal in memory and follow that rule. By contrast, our model uses domain-independent processes (structural alignment and similarity-based retrieval) to import information from the worked solution into the current problem.

The other transfer events are modeled incompletely. Our model does not have anything corresponding to part-of-line transfers. In any case, these events are extremely rare, accounting for just 3% of all transfer events in the protocols.

Klenk, M. and Forbus, K. 2007. Cognitive modeling of analogy events in physics problem solving from examples. In *The Proceedings of CogSci-07*. Nashville, TN.

The closest analog to search control is deciding to use a retrieved equation. Checking in our model corresponds to using a sanity check. Transfer failures are indicated when a precondition test fails, blocking the use of a candidate inference.

Given these differences, a quantitative comparison of the number of different analogy events would not be informative. However, in addition to the consistencies noted above, we would expect there to be a qualitative pattern of consistency. That is, we would expect that our model will predict more line transfer events than observed in protocols, since it is at the extreme poverty end of assumptions about initial knowledge. Second, analogical search control is done implicitly in our system; therefore our system has no explicit search control events. Third, we believe our model to be incomplete in terms of modeling sanity checking and transfer failures; therefore we expect to see fewer such events than in protocols.

Table 2: Analogy Events per Problem

Event Type	Companion's Problems			Protocols (n=24)
	Correct (n=64)	Failed (n=36)	Total (n=90)	
Initialization	1.12	1.16	1.13	1.2
Transfer	6.72	9.39	6.72	4.9
Line	6.13 (91%)	9.33 (99%)	7.18 (95%)	2.6 (54%)
Checking	.47 (7%)	.03 (<1%)	.33 (4%)	.75 (15%)
Failure	.12 (2%)	.03 (<1%)	.09 (1%)	.38 (8%)
Other	N/A	N/A	N/A	1.17 (24%)

Table 2 illustrates the number of analogy events by type per problem. The "Other" event type contains the analogy events not modeled by our system, including search control and part of line transfer as well as the miscellaneous events from the protocols. The Companion's analogy events are further divided depending on if the problem in which the event occurred was solved correctly. Also, percentages of total transfer events are supplied next to each transfer event in each condition.

Our results have a reasonable qualitative fit with VanLehn's human protocols. VanLehn summarized the results concerning type of information transferred in the protocols by saying "The basic result is simply that most students, both Good and Poor, transferred whole lines from the example to the problem" (1998, p. 364). Our system was even more dependent on line transfers for problem solving due to the fact that the vast majority of the domain knowledge used to solve these problems came from examples. The majority of the "Other" transfers were search control events, which motivates us to focus on search control events in our future work. While our model's sanity checking and failed transfer events are incomplete, both of these occurred more frequently on correctly solved problems. This, in addition to the fact that the protocols noted even more of these types of events per problem, indicates that a more complete model of these analogy events could lead to more robust problem solving.

Related Work

A number of prior analogical problem solving models, including PHINEAS (Falkenhainer 1990), Melis and Whittle's inductive theorem prover (1999), and Klenk *et al.*'s everyday physical reasoning (2005), import and adapt the entire solution of the example to the current problem. Other systems such as Cascade (VanLehn 1998) and ACT-R (Anderson 1993) use analogy primarily to overcome impasses in rule-based reasoning. Finally, there are systems such as Eureka (Jones 1989), derivational analogy in Prodigy (Veloso & Carbonell 1993), and APSS (Ouyang & Forbus 2006) that use analogy to improve efficiency.

Discussion and Future Work

We have shown how the Companions architecture can be used to model the use of analogy with worked solutions to solve AP Physics problems. We do not know of any other model which has been subjected to this kind of independent evaluation over this range of systematic variations in problem types, as well as having a reasonable qualitative fit to human protocol data. We find this very encouraging. However, the material we have tested it on only represents roughly 20% of the material in the Newtonian Mechanics portion of the AP Physics exam. Our goal is to expand the model where it can learn all of the material needed to solve AP Physics exams.

Three investigations are planned to achieve this goal. (1) Generalization from multiple analogies is an important aspect of human problem solving and in transitioning from novice to expert (Elio & Scharf 1990; Rieman *et al.* 1993; Kotovsky & Gentner, 1996). Therefore, we plan to construct generalizations using on SEQL (Kuehne *et al.* 2000), a structure mapping account for generalization and categorization, to facilitate the system's ability to apply its knowledge more broadly. For example, equations might be learned as encapsulated histories (Forbus 1984), which being parameterized could be used to model first principles reasoning. (2) As our system gains more domain knowledge, it will be necessary to extend our model to include analogical search control events. For this, we plan on incorporating the analogical search control mechanism used in APSS (Ouyang & Forbus 2006). (3) As Chi *et al.* (1981) note, one difference between novices and experts appears to be their encoding strategies. Consequently, we plan to explore methods for learning new encoding strategies, to capture this ability to move more directly from the everyday world to models that can be used to solve problems.

Acknowledgments

This research was supported by DARPA. We thank Cynthia Matuszek, Blake Shepard, and Casey McGinnis for

Klenk, M. and Forbus, K. 2007. Cognitive modeling of analogy events in physics problem solving from examples. In *The Proceedings of CogSci-07*. Nashville, TN.

their work representing the problems and the worked solutions. We thank Patrick Kyllonen, Catherine Trapani, and Vincent Weng at ETS for providing the testing materials and administering this experiment.

References

Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 153-166.

Chi, M.T.H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, 6, 271-315.

Chi, M.T.H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.

Chi, M.T.H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science***,** 15, 145-182.

Elio, R., & Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem solving strategy and knowledge organization. *Cognitive Science*, 14,579-639.

Faries, J. M., & Reiser, B. J. (1988). Access and use of previous solutions in a problem solving situation. In V. L. Pate1 & G. L. Groen (Ed.), *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Falkenhainer, B. (1990). A unified approach to explanation and theory formation. In J. Shrager and P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. Morgan Kaufmann Publishers.

Falkenhainer, B., Forbus, K. and Gentner, D. (1989). The Structure-Mapping Engine. *Artificial Intelligence*, (41).

Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence* 24:85–168.

Forbus, K. (2001). Exploring analogy in the large. In Gentner, D., Holyoak, K., & Kokinov, B. (Eds.) *Analogy: Perspectives from Cognitive Science.* MIT Press.

Forbus, K. & de Kleer, J. (1993). *Building Problem Solvers*, MIT Press.

Forbus, K. & Gentner, D. (1997). Qualitative mental models: Simulations or memories? *Proceedings of the Eleventh International Workshop on Qualitative Reasoning*, Cortona, Italy, June 3-6, pp. 97-104.

Forbus, K., Gentner, D., & Law, K. (1994) MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19, 141-205.

Forbus, K. & Hinrichs, T. (2006). Companion cognitive systems: A step towards Human-Level AI. *AI Magazine*, 27(2): pp 83-95.

Forbus, K. & Oblinger, D. (1990). Making SME greedy and pragmatic. In *Proceedings of CogSci-1990.*

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

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 99-129). Hillsdale, NJ: Erlbaum.

Jones, R. (1989). A model of retrieval in problem solving. Unpublished Doctoral Dissertation, University of California at Irvine, Irvine, CA.

Holyoak. K.J. (1985). The pragmatics of analogical transfer. In G.H. Bower (Ed.), *The psychology of teaming and motivation.* (19). New York: Academic.

Klenk, M., Forbus, K., Tomai, E., Kim,H., and Kyckelhahn, B. 2005. Solving everyday physical reasoning problems by analogy using sketches. *Proceedings of 20th National Conference on Artificial Intelligence (AAAI-05)*, Pittsburgh, PA.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. Child Development, 67, 2797-2822.

Novick, L.R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14.

Ouyang, T. and Forbus, K. 2006. Strategy variations in analogical problem solving. *Proc. of AAAI-06*

Reimann, P., Wichmann, S., & Schult, T. J. (1993). A learning strategy model for worked-out examples. In P. Brna. S. Ohlsson, & H. Pain (Eds.), Artificial *Intelligence in Education: Proceedings of AI-ED93* (pp. 290-297). Charlottesville, VA: Association of Advancement of Computing in Education.

Ross, B. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629-639.

Ross, B. (1989). Remindings in learning and instruction. In S. Vosniadou & A. Ortony (Ed.), *Similarity und analogical reasoning*. Cambridge: Cambridge University Press.

VanLehn, K., Jones, R. M., & Chi, M.T.H. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, 2(1).

VanLehn, K., & Jones, R. M. (1993a). Learning by explaining examples to oneself: A computational model. In S. Chipman & A. Meyrowitz (eds.), *Cognitive Models of Complex Learning*. Boston, MA: Kluwer Academic Publishers.

VanLehn, K., & Jones, R. M. (1993b). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polson (Ed.), *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

VanLehn, K. (1998). Analogy events: How examples are used during problem solving. *Cognitive Science* 22(19): 347-388.

Veloso, M. and Carbonell, J. (1993). Derivational analogy in PRODIGY: Automating case acquisition, storage, and utilization. *Machine Learning*, 10:249-278.

Klenk, M. and Forbus, K. 2007. Cognitive modeling of analogy events in physics problem solving from examples. In *The Proceedings of CogSci-07*. Nashville, TN.