Similarity-based Cognitive Architecture

Kenneth D. Forbus and Dedre Gentner
Institute for the Learning Sciences
Northwestern University
1890 Maple Avenue
Evanston, Illinois, 60201, USA

Similarity permeates human cognition. There is evidence that objects are categorized based partly on similarity to previous category members and that the likelihood of transfer is governed by the similarity between the original and current situations. New problems are often solved by analogy to prior problems. Similarity is responsible for many human errors, such as perceptual confusions and many recall intrusions; but at the same time, analogy and similarity are important in scientific discovery. Consequently, we are developing a cognitive architecture in which similarity computations play a central role. This is unlike most architectural approaches, which either do not treat analogy and similarity at all, or relegate them to a subsidiary role, to be called in sporadically when other mechanisms are stuck. We are using Gentner's Structure-Mapping theory [12] as our framework for defining similarity computations.

The rest of this note addresses the list of issues suggested by the symposium organizers.

1 Why integration?

Part of our motivation is the long-range goal of creating a computational account of human reasoning and learning in physical domains [9]. That is, we are trying to capture the processes and representations that make it possible for someone to learn about areas such as thermodynamics from observation, experimentation, and instruction. We want to capture the entire progression of human mental models, from the accumulation of prototypical behaviors through the ability to perform engineering analyses, as well as the computations which move a learner from one model to another. Most of our efforts to date have been short forays into specific subproblems. While much can be learned this way, and such efforts will continue, some issues can only be addressed by looking at larger pieces of the problem. Here are two specific projects to illustrate what we mean:

Learning from lay-science texts: Many introductory science books focus on imparting qualitative knowledge about a domain, providing more systematic explanations for phenomena that the reader may have already observed and linking it to new phenomena. Such books typically use analogy to convey models, and often build up a domain model by multiple, interacting analogies. Our goal is to construct a program which can build up a qualitative model of a domain from such texts that will enable it to answer questions it couldn't before.

Learning engineering thermodynamics: Thermodynamics is a substantially harder domain than those traditionally used in problem-solving studies: The collection of techniques which suffice for puzzles and even for a subset of newtonian mechanics are woefully inadequate to capture human performance in this domain! The goal of this project is to build a system which can learn to perform on engineering thermodynamics problems as well as a college student after taking an introductory course. We presume the system starts with good (albeit partial) qualitative models. Learning will proceed by processing textbook information and attempting to solve new problems posed by an instructor. Our focus here is on modeling the acquistion of quantitative knowledge and problem-solving skills, which includes the effective integration of such knowledge with the system's intuitions (as represented by its qualitative model). For instance, we want the system to be able to absorb and integrate information from multiple sources, including diagrams. Another issue we want to study is how to design the system to profitably take advice from someone who doesn't know its detailed internal state. Quite apart from cognitive modeling, as our knowledge bases grow, such techniques will become crucial for augmenting and even maintaining them.

Both experiments involve integrating problem-solving, analogical learning, knowledge representation, spatial reasoning, memory, and (to some degree) natural language.

1.1 Basic components

To date we have concentrated on developing accounts of the basic components required, with forays into particular aspects of the problem. These forays include Falkenhainer's PHINEAS [7] program, which explored learning at the Naive Physics stage, and G. Skorstad's SCHISM [24] which is exploring how to integrate qualitative and quantitative models to solve engineering thermodynamics problems. PHINEAS included the following components:

- SME [6], a simulation of structure-mapping.
- QPE [8], an envisioner for Qualitative Process theory.
- ATMoSphere, an ATMS-based inference engine with antecedent rules and an and/or graph control system.
- DATMI [4], a measurement interpretation system.
- TPLAN [16], an Allen/Koomen-style temporal planner.

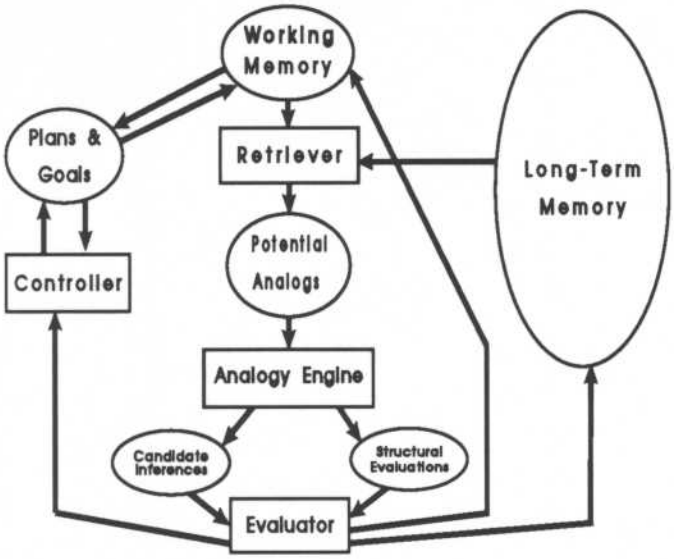
while SCHISM currently uses only QPE and ATMoSphere.

Much of our research effort has involved building and extending these components. For example, we recently extended SME to make it more efficient and more suitable as a component in problem solvers [11]. An important property of those systems intended to be cognitive simulations is what we call *accountability*. That is, processing choices not explicitly constrained by theory must be easily changable, so that the dependence of results on alternate choices can be explored. For instance, SME's input includes two sets of rules which construct and evaluate local matches, allowing it to be programmed to emulate all the comparisons of structure-mapping and other matching theories consistent with its basic assumptions [5].

Smaller combinations of these systems have been used to model particular aspects of cognitive processes. For example, J. Skorstad's SEQL, which uses SME as a module, provides a toolkit for exploring exemplar-based versus abstraction-based models of concept formation. SEQL has been used to model

Figure 1: The Structure-Mapping Architecture

This diagram illustrates how human analogical processing may be organized. Can this organization be extended to cover a broad range of cognition? We intend to build a series of simulations to explore the characteristics of similarity-based architectures



data concerning sequence learning effects in geometric stimuli [26].

Another example is MAC/FAC [13,14], an initial exploration of similarity-based retrieval and inference. Psychological results indicate that human similarity-based retrieval from long-term memory is largely driven by surface commonalities; while in contrast, human judgements of similarity and inferential soundness are chiefly driven by the degree of structural match [15,17,22]. MAC/FAC, for "many are called, but few are chosen", is a two-stage retrieval system that attempts to capture the different roles of similarity in this phenomena. The first stage (MAC) is a computationally cheap, but structurally stupid match process. Given a probe, MAC selects a subset of memory for further processing using a *numerosity match*, a coarse, non-structural means of estimating the quality of a structural match. Thus while some of the matches it returns are sound, many of them need not be. The FAC stage applies the full structure-mapping match computation, which means fully enforcing structural consistency, producing global interpretations, and calculating candidate inferences (i.e., the surmises which the match suggests). The FAC stage currently consists of SME operating in literal similarity mode (i.e., sensitive to both structural and object-based similarity).

Figure 1 illustrates the design of our architecture. In our current version, the Retriever and Analogy Engine are subsumed by MAC/FAC. We plan to experiment with several organizations of the Working Memory, the Evaluator, and the Controller. The questions we want to explore include:

1. What instantiations of these modules suffice to provide at least the power of traditional AI problem-solvers, but are consistent with psychological data? How much of the work can be borne by similarity computations?
2. How does analogy interact with more traditional problem-solver organizations? When should a problem-solver resort explicitly to analogy, and how can implicit learning be integrated with problem-solving?

To explore these questions we plan to build a series of inference engines. Each engine will attempt to perform more and more of the inferential work by similarity computations. For example, the first engine will use a pattern-directed rule system with an underlying truth-maintenance system to perform most of the reasoning, with SME used to generate surmises about solutions based on hints (e.g., "Look at this previously-solved problem"). Next, the pattern-directed rule system could be replaced by MAC/FAC. A SEQL-like system could then be integrated to provide a model of implicit learning, abstracting commonalities from frequently encountered classes of situations to model the process of rules naturally arising from cases as expertise increases in a domain.

Analogical learning from lay-science texts requires tapping into a broad understanding of the world. Therefore we will be attempting to build on the CYC knowledge base [19], extending its ontology with the constructs of Qualitative Process theory and interfacing it with SME.

2 Sources of inspiration

The SOAR project and Van Lehn's SIERRA have been major sources of inspiration, along with work in case-based reasoning [18] and instance-based models of human memory (e.g., [20]). Collaborations with Doug Medin and Jerry DeJong have sparked our interests in exploring categorization and problem solving, respectively.

3 Characterization

Our answers to the dimensional decomposition suggested by the organizers are based on the proposed learning projects described above.

3.1 Generality

Many of the specific representational content and techniques (e.g., languages for physical processes and mathematics) should be applicable to a broad range of scientific and technical domains. We hope that at least a subset of our mechanisms will prove to be valid models for human cognition, independent of domain.

We are planning experiments with other kinds of domains as well. It is worth looking at other problem-solving domains, for instance, to better compare our ideas with other systems. One interesting example might be geometry theorem proving (c.f. [1]). However, we see an important line of research being the simulation of developmental data. Developmental psychology is currently making great strides in characterizing children's mental models. One example is the exploration of causal understanding of motion and collisions in infants [2]. Another is the development of theories of weight and balance [3,23]. We want to develop psychologically plausible computational accounts of these models and their acquisition.

3.2 Versatility

Learning by experience in the world is an important aspect of human learning in physical domains. However, we have no plans for integrating real vision or real robotics. We view these areas as extremely difficult research problems in their

own right. Our suspicion is that, given the current state of the art and our goals, integrating in that direction would be unproductive for us, unless it was in collaboration with other researchers whose research focus was vision and/or robotics. Our approach instead is to concentrate on the spatial reasoning problems that arise in physical reasoning and ignore taking action in the physical world[1].

3.3 Rationality

It would be a fairly poor cognitive model if its actions were always consistent with its knowledge and goals, wouldn't it?

3.4 Ability to add new knowledge

We are hoping to move from "mind implants" to something more akin to instruction, where there is a teacher who has only a glimmer of the system's internal state, based on observing it. This is one reason for scaling up: In today's knowledge-poor simulations, it is altogether too easy to make very detailed predictions about the internal state of the system by a few observations because they simply can't do very much.

3.5 Ability to learn

Presumably.

3.6 Taskability

One way to view this question is, "Are we trying to develop computer individuals, as Nilsson suggested in 1983?" The short answer is: not yet. One common theme in the projects above that echos Nilsson's suggestion (and a difference from most learning programs) is that we want to build programs whose knowledge bases evolve over a significant span of learning experiences – i.e., working through a textbook. We still know very little about building robust systems which can survive such extended bouts of operation and learning[2] However, for the forseeable future we still intend to tell our programs what to do, at least in broad terms.

3.7 Scalability

We certainly hope so. But we expect there will be problems.

3.8 Reactivity

In the quiet world of book learning, we suspect most of the system's surprises will be conceptual boggles rather than sudden environmental threats.

3.9 Efficiency

Implementing detailed simulations of cognitive mechanisms on today's hardware can be very difficult. It seems likely, for instance, that massive parallelism will be necessary for modeling some aspects of human memory phenomena. For some experiments we focus on getting the model "right" and damn the actual run-times. But for these projects, our challenge is to find good approximations to cognitively plausible mechanisms that will allow us to explore issues at the larger scale.

3.10 Psychological validity

There are a number of questions we want to explore computationally which have seen little attention in previous cognitive architecture studies. These include issues concerning the form and role of different kinds of knowledge about the physical world as well as a detailed explication of the role of similarity computations in cognition.

But there are also several classic issues which are unavoidable. One of them is memory organization. AI models of memory, which tend to be based on clever indexing schemes, seem unlikely to scale to human-size knowledge bases and the demands of significant conceptual change during learning. Psychological models of memory organization have typically utilized very simple representations, such as feature vectors. Feature vectors can provide tractable large-scale searches, but they fail to capture the rich relational information that people clearly possess and use in reasoning. On the other hand, matching structured descriptions tends to be computationally expensive, making large searches seem unfeasible. MAC/FAC's two-stage computation provides the best of both worlds: The search carried out by the MAC stage is, in effect, based on a flat, simple representation. While not highly accurate [13], its low computational cost makes large-scale searches feasible. The full, structured representations it retrieves are then used by the FAC stage, thus providing the relational matches required to draw inferences. This allows the MAC/FAC model to capture two seemingly incompatible intuitions about memory: Access tends to be governed by surface properties, while inference tends to be governed by relational matches. We think that by looking at reasoning and learning in a complex domain we may gain new insights about how memory works.

4 Knowledge sharing

One problem with PHINEAS and SCHISM was that a substantial portion of each system's expertise was frozen in impenetrable rule mechanisms. Some impenetrability is probably okay; we presume that the laws of qualitative mathematics are already known, for instance, and hence do not have to be learned. But we need to make our next generation programs more transparent. We are hoping CYC will help in this.

5 Control

We have done very little thinking about this. Suggestions are welcomed.

6 Comparison with other cognitive architectures

We agree with Newell [21] that the field should be exploring a variety of architectural approaches. We find much that is exciting and admirable in the SOAR, ACT*, and SIERRA projects. However, we differ from them in three important ways.

First, we assume that important general constraints on architectures will come from a better understanding of the representational needs imposed by rich domains and human-quality robustness and performance. Most architectural studies have focused on simple domains and small knowledge bases. We believe many of the distinctions which separate today's brittle AI systems from the quality of human cognition

[1] We think of metric diagrams and visual routine processors as the same thing, viewed from different sides of the cognition/perception borderland.

[2] The CYC project is the only effort we know of which has faced at least some aspects of this problem squarely.

can only be understood by looking at reasoning and learning in complex domains.

For example, we conjectured that naturalistic representations would include a preponderance of appearance and low-order information, unlike current AI representations which tend to focus on task-relevant information (the *specificity conjecture*). This conjecture allowed us to constrain the space of possible algorithms for structural evaluation of analogies [10].

Second, like Van Lehn, we consider content-oriented psychological evidence to be a crucial source of constraint. Process-oriented measures, such as reaction time studies and numerical measures of human performance can provide valuable information once there is an overall framework to ground their interpretation. However, we believe content-oriented evidence, such as patterns of recall and classification, protocol studies, and assessment of mental models, will be crucial to arriving at the correct overall framework. To us, the critical tests include the ability to assimilate new information about a domain from a lay-science text, and to parlay this understanding with additional instruction into the ability to solve new problems in the domain.

Third, we believe that a similarity-based architecture will ultimately provide a more constrained account of cognitive processes than production-rule systems. Production rules provide little constraint on the representation of knowledge, since there are many equivalent ways to encode a particular computation in them. Without such constraints it is difficult to make detailed predictions using these theories, since a change in representation could yield substantially different results. In Structure-Mapping, analogy and similarity computations are sensitive to the form of the representation. This sensitivity means that our representations should be less tailorable than standard production-rule models – not only must they carry out the required inferences, but they also must perform reasonably under similarity computations. Whether or not this extra constraint leads to additional discriminability is, of course, an empirical question.

7 Acknowledgements

We thank Art Markman, Gregg Collins, and Larry Birnbaum for helpful suggestions. This research is funded by the Office of Naval Research, through the Cognitive Sciences Division and the Computer Science Division.

References

[1] Anderson, J., Greeno, J. Kline, P., and Neves, D. "Acquisition of problem-solving skill", in Anderson, J. (Ed.), *Cognitive Skills and Their Acquisition*, Lawrence Erlbaum Associates, Inc., 1981.

[2] Baillargeon, R. "Young infants' reasoning about the physical and spatial properties of a hidden object", *Cognitive Development*, **2**, pp 179-200, 1987.

[3] Carey, S. *Conceptual Change in Childhood*, Bradford Books, MIT Press, Cambridge, MA, 1985.

[4] DeCoste, D. "Dynamic Across-Time Measurement Interpretation", Proceedings of AAAI-90, Boston, MA, 1990.

[5] Falkenhainer, B., The SME user's manual, Technical Report UIUCDCS-R-88-1421, Department of Computer Science, University of Illinois, 1988.

[6] Falkenhainer, B., Forbus, K., Gentner, D. "The Structure-Mapping Engine: Algorithm and examples" *Artificial Intelligence*, 41, 1989, pp 1-63.

[7] Falkenhainer, B. "Learning from physical analogies: A study in analogy and the explanation process", Ph.D. thesis, University of Illinois, Urbana. 1988.

[8] Forbus, K. "QPE: A study in assumption-based truth maintenance" *International Journal of Artificial Intelligence in Engineering*, October, 1988.

[9] Forbus, K. and Gentner, D. "Learning Physical Domains: Towards a theoretical framework", Proceeding of the 1983 International Machine Learning Workshop, Monticello, Illinois, June, 1983. (An expanded version appears in Michalski, R., Carbonell, J. and Mitchell, T. "Machine Learning: An Artificial Intelligence Approach, Volume 2", Tioga press, 1986.)

[10] Forbus, K. and Gentner, D. "Structural evaluation of analogies: What Counts?", Proceedings of the Cognitive Science Society, August, 1989

[11] Forbus, K. and Oblinger, D. "Making SME greedy and pragmatic", Proceedings of the Cognitive Science Society, July, 1990.

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

[13] Gentner, D. "Finding the needle: Accessing and reasoning from prior cases" Proceedings: Case-based reasoning workshop, DARPA ISTO, Pensacola, Florida, May, 1989.

[14] Gentner, D. and Forbus, K. "MAC/FAC: A Model of Similarity-based Access and Mapping", Proceedings of the Cognitive Science Society, to appear, August, 1991.

[15] Gentner, D. and Landers, R. "Analogical reminding: A good match is hard to find", Proceedings of the International Conference on Systems, Man, and Cybernetics, (pp 607-613), Tucson, Arizona, 1985.

[16] Hogge, J. "TPLAN: A temporal interval-based planner with novel extensions", Technical Report No. UIUCDCS-R-87-1367, September, 1987.

[17] Holyoak, K.J. and Koh, K. "Surface and structural similarity in analogical transfer", *Memory and Cognition*, **15**, pps 332-340, 1987.

[18] Kolodner, J.L. (Ed.) *Proceedings of the First Case-Based Reasoning Workshop*, Morgan Kaufmann, Los Altos, CA, 1988.

[19] Lenat, D. B. and Guha, R.V. *Building Large Knowledge-Based Systems*, Addison-Wesley, Reading, MA, 1990.

[20] Medin, D.L. and Schaffer, M.M. "Context theory of classification learning", *Psychological Review*, **85**, pp 207-238, 1978.

[21] Newell, A. "Towards unified theories of cognition", unpublished manuscript.

[22] Ross, B.H. "Remindings in learning and instruction", in S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning*, Cambridge University Press, London, pp 438-469, 1989.

[23] Seigler, R.S. "Mechanisms of cognitive growth: Variation and selection." In R.J. Sternberg (Ed.), *Mechanisms of Cognitive Development*, Waveland Press, Prospect Heights, IL, 1988.

[24] Skorstad, G. and Forbus, K. "Qualitative and quantitative reasoning about thermodynamics", Proceedings of the Cognitive Science Society, August, 1989.

[25] Skorstad, J., Falkenhainer, B., and Gentner, D. "Analogical processing: A simulation and empirical corroboration", Proceedings of AAAI-87.

[26] Skorstad, J. Gentner, D. and Medin, D., "Abstraction processes during concept learning: A structural view", Proceedings of the Cognitive Science Society, 1988.