Title
Qualitative Modeling and Similarity in Back of the Envelope Reasoning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 25(25)

ISSN
1069-7977

Authors
Paritosh, Praveen K.
Forbus, Kenneth D.

Publication Date
2003

Peer reviewed

Qualitative Modeling and Similarity in Back of the Envelope Reasoning

Praveen K. Paritosh (paritosh@cs.northwestern.edu)
Kenneth D. Forbus (forbus@northwestern.edu)
Qualitative Reasoning Group, Department of Computer Science,
Northwestern University, 1890 Maple Ave,
Evanston, IL 60201 USA

Abstract

Back of the envelope reasoning involves generating quantitative answers in situations where exact data and models are unavailable and where available data is often incomplete and/or inconsistent. A rough estimate generated quickly is more valuable and useful than a detailed analysis, which might be unnecessary, impractical, or impossible because the situation does not provide enough time, information, or other resources to perform one. Such reasoning is a key component of commonsense reasoning about everyday physical situations. This paper presents a similarity-based approach to such reasoning. In a new scenario or problem, retrieving a similar example from experience, sets the stage for solving the new problem by borrowing relevant modeling assumptions and reasonable values for parameters. We believe that this tight interweaving of qualitative and analogical reasoning is characteristic of common sense reasoning more broadly. Understanding the *feel* for magnitudes is another crucial aspect of such reasoning, and incorporating effects of quantitative dimensions in similarity judgments and generalizations, hitherto unexplored, raises very interesting questions.

1 Introduction

We live in a world of quantitative dimensions, and reasonably accurate estimation of quantitative values is necessary for understanding and interacting with the world. Our life is full of evaluations and rough estimates of all sorts. How long will it take to get there? Do I have enough money with me? How much of the load can I carry at once? These everyday, common sense estimates utilize our ability to draw a quantitative sense of world from our experiences.

Back of the envelope (BotE) analysis involves the estimation of rough but quantitative answers to questions where the models and the data might be incomplete. In domains like engineering, design, or experimental science, one often comes across situations where a rough answer generated quickly is more valuable than waiting for more information or resources. Some domains like environmental science [Harte, 1988] and biophysics [O'Connor and Spotila, 1992] are so complex that BotE analysis is the best that can be done with the available knowledge and data. BotE reasoning is ubiquitous in daily life as well. Common sense reasoning often hinges upon the ability to rapidly make approximate estimates that are fine-grained enough for the task at hand. We believe that the same processes underlie both these common sense estimates and expert's BotE reasoning to generate ballpark estimates. Specifically, the drawing upon experience to make such estimates, and the achievement of expertise in part by accumulating, organizing, and abstracting from experience to provide the background for such estimates, are the same fundamental processes. We claim that qualitative reasoning [Forbus, 1984] is essential for such analyses for two reasons:

1. *Qualitative models provide analytic framework.* Understanding what entities and physical processes are relevant is crucial in determining what parameters are relevant. Modeling assumptions expressed in terms of the conceptual understanding of the situation determine when particular quantitative estimation techniques are appropriate.
2. *Qualitative models facilitate comparison*. Similarity in qualitative, causal structure helps determine what experience is relevant when making an estimate. Similarity is also used in helping evaluate the reasonableness of an estimate. Including qualitative descriptions in remembered experiences along with quantitative data facilitates comparison and abstraction from experiences.

A combination of QR and experiential knowledge seems to be the key to BotE reasoning. QR helps us determine what phenomena are relevant, and experiential knowledge supplies useful default and pre-computed information, including both numeric values and relevant modeling assumptions, as well as knowledge about similar situations that can serve as a reality check for the estimates. The need to compare parameters and to make estimates guided by similarity in turn raises interesting questions about what role(s) quantitative dimensions play in our judgments of similarity, and how we develop our quantitative sense of a domain with experience.

In this paper, we look at quantitative estimation (also called rough estimation, back of the envelope analysis, etc), which we believe highlights some of the very important questions at the intersection of analogical and qualitative reasoning [Forbus and Gentner, 1997], and more. We argue that BotE reasoning provides a fertile ground for exploring key aspects of common sense reasoning, and present our approach towards modeling it. Section 2 presents a brief review of relevant research. Section 3 elaborates on our approach. Section 4 contains two extended examples that illustrate our arguments. Section 5 presents some open research issues and our initial attempts to address them, and we wind up with our plans for future work.

2 Related Work

This section is divided into three subsections. We start with a review of psychological work on real-world quantitative estimation of dimensions and probabilities. In section 2.2, we review models of similarity. Section 2.3 differentiates our work from semi-quantitative reasoning.

2.1 Psychology of Quantitative Estimation

Peterson and Beach (1967) review a set of psychological studies to test people's abilities to derive statistical measures of populations and samples such as proportions, means, variances, correlations, etc. Although some of the studies have conflicting results, the key result that people are quite good at abstracting measures of central tendency, and there are systematic differences in intuitive judgments and objective statistical values. For example, people don't weigh all deviations equally in computing variance. Instead, they are quick to believe in a distribution even from a few samples, and tend to be conservative in revising their measures on the basis of new data points. Tversky and Kahneman (1974) reported people's assessment of probabilities of uncertain events. In a very important set of results, they show that people make systematic errors because of a set of heuristics that they employ.

Brown and Siegler (1993) proposed a framework for real-world quantitative estimation called the *metrics and mappings* framework. They make a distinction between the quantitative, or metric knowledge (which includes distributional properties of parameters), and ordinal information (mapping knowledge). Through a set of experiments they showed that the ways people revise and assimilate quantitative and ordinal information are quite different. Their experiments involved subjects making quantitative estimates of populations of ninety-nine countries. Afterwards participants were told the correct value for populations of 24 of the countries, and then they went through and re-estimated the full set of 99 populations (the 24 seed countries and 75 transfer countries). Metric properties (as measured by sum of absolute value of errors for all of the estimates) improved, but ordinal knowledge (the order of different population, as measured by the rank-order correlation) remained unchanged. On the other hand, telling them laws like "Population of European countries are generally overestimated", and "Population of Asian countries are generally underestimated", improved their ordinal knowledge.

Linder (1999) studied quantitative estimation in the context of engineering education. Based on responses to real world questions, he tried to build a framework for how people do rough estimations. About a hundred mechanical engineering seniors at MIT, and fifty each at five other universities attempted these estimation questions. He also compiled responses from a hundred professionals, out of which about there were about thirty each of electrical and mechanical engineers, and the rest from other engineering and science backgrounds. His focus was how to improve engineering curricula, and thus his framework is informal and not couched in computational terms; nevertheless, it provides an interesting source of data. In one experiment, when people were asked to estimate dimensions of an aluminum bar, more than 50% came up with correct estimates and all the answers were in the correct order of magnitude. However, in the same experiment, when people were asked to estimate the power of a DC motor, only about 30% got it right and the responses varied by six orders of magnitude! We suggest a possible explanation for this discrepancy in terms of our model below.

2.2 Models of Similarity

In the 1960s, a popular psychological model for similarity was to represent objects as points in a psychological space of stimulus dimensions, where similarity is defined as the distance between points. Multidimensional scaling [Shepard, 1962] is a technique designed to uncover this psychological space by analyzing people's similarity judgments. This work drew a distinction between integral and separable dimensions, and explored how this distinction affects our similarity judgments. Tversky's set-theoretic account (1977), where feature commonalities and feature differences both affect the similarity between two concepts, raised many questions about the metric space model. Gentner's (1983) structure-mapping theory provides an account of analogy and similarity that better fits the psychological data than either feature space or feature set models. For example, structure-mapping handles relationships as well as features, which is crucial for the use of similarity in reasoning. The idea of structural alignment also provides deeper insights into the comparison process that has led to many new predictions. For example, Markman and Gentner (1993) proposed a structure-based model that makes three distinctions: commonalities, alignable differences and non-alignable differences. Alignable differences are differences along the same roles in two representations, whereas non-alignable are differences along different roles. So, a hotel and motel have a lot of alignable differences, whereas a hotel and motorbike has a lot of non-alignable differences. In their recent studies, they have shown that people value alignable differences more than non-alignable while making similarity judgments.

2.3 Semi-quantitative Reasoning

It is important to distinguish between the notion of quantitativeness in semi-quantitative reasoning [Berleant and Kuipers, 1997] and BotE reasoning. In semi-quantitative reasoning, functional uncertainty is represented by defining envelopes within which functional constraints must lie, and parametric uncertainty is represented by numeric intervals. Clearly, this is still in the spirit of purely first-principles reasoning, in contrast to our similarity-based approach to model formulation and parameter estimation.

3 A Similarity-Based Model of BotE Reasoning

Back of the envelope reasoning involves the estimation of rough but quantitative answers to questions. Most of the questions are real-world problems, where usually one does not have complete or accurate models or model parameters. Yet one can get a lot out of approximate estimates. This type of reasoning is particularly common in engineering practice and experimental sciences, including activities like evaluating the feasibility of an idea, planning experiments, sizing components, and setting up and double-checking detailed analyses. There is a tradeoff between specificity (resolution and certainty in the answer) and economy. As we try to increase the specificity in the answer, the analysis requires more resources in the form of time, information, formalization, and computation; and one might not have one or more of these at hand. There is a large variety of such questions, such as

Q1. Estimate the amount of work a person does shoveling the walk after a snowstorm.

Q2. Estimate the drag force on a bicycle and rider traveling at 20 mph.

Q3. Estimate the energy stored in a new 9-volt transistor battery.

Q4. Estimate the tension of a car's safety belt if the car crashes into a pillar (at speed of 30km/h and produces a 30 cm deep dent).

Q5. How long does it take to reach home from your office, or to get ready in the morning?

Q6. How much money would you be spending on that vacation you have planned?

Q7. You know a recipe that you made for yourself some time back – now you have to make it for eight people, and you want it less spicy and you ran out of one of the ingredients.

Questions 1 to 4 are questions that might arise in engineering circumstances, whereas Questions 5 to 7 are questions that arise in daily life. Question 5 seems more based on direct observation than others. For example, you might have earlier noticed how much time it takes for you to arrive, or what were your best/worst times, and you recall those, and might employ some measure of central tendency to come up with a time estimate. In Question 6 (and others), it seems that one must build a simple *estimation model*, and use this model to answer the question by estimating in turn values for the parameters in the model.

Essentially, BotE reasoning involves coming up with a numeric estimate[1] for a parameter. This can be decomposed into two distinct (but not independent) processes.

Direct parameter estimation – This involves directly estimating a parameter based on previous experience or domain knowledge. For instance, we might know the value of a physical constant, or use a value from a previous example that is highly similar to the current problem, or combine multiple similar examples to estimate a value based on those prior values. Or, it might be that with enough experiences in a domain, one has developed a *feel* for magnitudes. The knowledge and processes involved in developing that, for concreteness, we'll refer to as the *sense of the quantitative*, and in Section 5, we outline our hypotheses about how that comes about to be.

Building an estimation model – This is required when the parameter to be estimated is not usually directly stored or encountered. In such cases one has to build a model that relates the parameter in question to other parameters, which in turn must be estimated.

Lets look at a small example to make this distinction clear. Consider the question – How many pieces of popcorn would fill the room you are now sitting in? The parameter, `num-popcorn` is not one that one can recall a value from the memory – so one way to derive it would be

```
num-popcorn = volume-room/volume-
         popcorn          …(1)
```

Approximating room to a cuboid, and popcorn to a cube (considering the voids left after packing in popcorn kernels[2] this is a reasonable assumption),

```
num-popcorn = l*b*h / a^3     …(2)
```

where `l`, `b`, `h` are length, breadth and height of the room and `a` is the edge of the cube that describes a popcorn. In (2), we have built an estimation-model for the number of popcorn kernels, which we have now described in terms of a set of parameters that can be estimated by direct parameter estimation. Estimation-model building can be recursive (after our initial model in (1), we had to build sub-models for the volumes of the room and popcorn).

What makes someone good at BotE reasoning? Experience with similar estimation tasks, ability to compare a parameter with other known values, ease of access to estimation models seem to be some of the important factors in numeric estimation skill. Some parameters are clearly more accessible than others, and there are strong domain expertise effects, too. One of the important things as one learns a domain is extensive familiarity with the quantitative aspects of a domain: when is a parameter value to be reasonable/typical, or high, or on the conservative side, etc. This could explain Linder's (1999) results about the variability in accuracy of BotE reasoning. It is not surprising that the intuitions of an electrical engineer about motors and batteries is more accurate than an a mechanical engineer's intuition, or that mechanical engineers' answers about drag force and tension are more accurate than those of electrical engineers. What is this experiential knowledge, and how exactly does that help in BotE?

1. Knowing a large number of examples of various problems and scenarios helps in building the estimation model. Given a new problem, we can solve it by retrieving a similar example from which we can borrow relevant modeling assumptions, default values, etc.
2. Exposure to a large number of examples involving various quantities in a domain gives rise to sense of the quantitative.

Thus we see analogical reasoning about within-domain experience as being central both to building estimation models and to selecting reasonable values for model parameters. To make these ideas clearer, we turn to some extended examples for illustration.

---

[1] We emphasize the numeric/quantitative aspect of such reasoning which is in no way in any conflict with our goals are to understand human qualitative reasoning. Qualitative should not be thought as necessarily being not quantitative!

[2] Of course, if we didn't have the volume of a popcorn in our domain theory, the fact that a cube is a reasonable approximation for a single popcorn and its related void, is an interesting (and general purpose) estimation modeling strategy.

4 Extended Examples

In this section we look at two examples that illustrate various points that we made earlier. Both the questions in this section were used in Linder's study.

Q2 Estimate the drag force on a bicycle and rider traveling at 20 mph (9 m/s).

One of the important things to note about this problem (which is the case with most of real-world estimation tasks) is that it is not completely specified. The basic description of the physical situation is very abstract, and most of the quantitative information that is needed to solve the problem is not provided. Several subjects, given this problem, indicated that they pictured a person on a bicycle from a distance from the side and/or the front; and often they made sketches of these views [Linder, 1999]. This strongly suggests to us that the model formulation phase itself involves retrieving a similar known scenario, to fill in the details.

Solution I This is a very simple solution. All of the power generated by the human is used up in propelling the bicycle at the given speed, and that all of it goes to overcoming the drag force. Since the estimate of the power that the human is producing while cycling under given conditions is the only parameter that it uses, the estimate strongly depends upon how representative the estimate of power is in the circumstances of the problem.

Table 1: Solution I for Q2

Model	Power = Force * Velocity
Parameters	Power (produced by the human during cycling) = 200 Watts Velocity (given) = 9m/s Force (to be estimated)
Solution	$F_{\mathbf{drag}} = 200/9 \approx 22$ N

In the direct parameter estimation for power, it is key that we look for human power output during similar activity. It turns out that humans can comfortably produce 100 watts of power, and up to 1500 watts in spurts.

Solution II This is the more standard solution that a mechanical engineer would come up with. The drag equation (1), which helps calculate the drag force on a moving object due to surrounding fluid, is definitely relevant to the problem. The difficulty though is that it has a bunch of other parameters that we don't know of, e.g., the drag coefficient, density of air, reference area of the body. The drag coefficient (C_{drag}) itself captures all complex dependencies (on the viscosity and compressibility of air, geometry of the body, and the inclination to flow) and is usually derived empirically. We look for similar scenarios, and indeed there is one, of human falling with terminal velocity (maybe in context of skydiving, and this is not a rare piece of information, considering that quite a few people did use this). In the free-fall scenario, the terminal velocity is known, and the drag force is known (as it counterbalances gravity, it equals the weight of the person). This allows us to estimate the constant of proportionality in the drag equation (2), and thus the drag force during cycling.

Table 2: Solution II for Q2

Model	$F_{\mathbf{drag}} = C_{\mathbf{drag}}\,(1/2\ \rho V^2)\ A$ …(1) Or, $F_{\mathbf{drag}} = KV^2$ for same sized objects in the same density fluid. …(2) Plugging the value of K back into (2) gives us $F_{\mathbf{drag}}$.
	Similar scenario: Free-fall, known terminal velocity, $V_T = 50$ m/s Here, $F_{\mathbf{drag\_free\_fall}}$ = Weight. …(3) $K = F_{\mathbf{drag\_free\_fall}} / V_{\mathbf{T}}^{2}$ = Weight/ $V_{\mathbf{T}}^{2}$ …(4)
Parameters	[A, $C_{\mathbf{drag}}$, ρ (density of air)] can be lumped into K, V (velocity), $V_{\mathbf{T}} = 50$ m/s
Calculations	K = 750/ 50^2 = 0.3 $F_{\mathbf{drag}} = 0.3 * 9 * 9 \approx 25$ N

Q3 Estimate the energy stored in a new 9-volt transistor battery.

This problem is an interesting example, where first principles reasoning from the chemistry of energy generation in the battery involves complicated domain knowledge, and none of the people asked even attempted to reason that way. What most of the people did was to imagine scenarios where such a battery was being used, and try to think from there. And the thing that is beautiful is the fact that this calculation gives us an estimate that is just as good as the more complex method. This is a nice example of where, for the purposes of BotE estimates, ability to successfully reason from known scenarios and examples buys us as much as far more first principles knowledge would. The solution below presents reasoning with very little knowledge about the battery. If I don't know anything about 9-volt battery, what is the next similar thing? A lot of people thought about car batteries, 1.5-volt AA batteries, etc.

This example also demonstrates that using examples allows us to transform the problem into ways that parameter estimation, or model building become more intuitive or accessible. For example, knowledge of parameters like the rated capacity of the battery, or, resistive load of the bulb would have led us to solutions, but we think in terms of parameters that are more accessible to us. Besides helping understand common sense qualitative reasoning, this is a great problem solving strategy for scientific and engineering reasoning as well.

Table 3: Solution for Q3

Model	Suppose I did not know anything about the 9v battery except its size, but I knew examples of where 1.5v AA batteries were being used. If I make the assumption that these two batteries are fundamentally the same, and only the difference in volume should be responsible for difference in energies stored. $E_{\mathbf{transistor}}/E_{\mathbf{AA}} = V_{\mathbf{transistor}}/V_{\mathbf{AA}}$ …(1)
	In a small hand-held flashlight, all the power provided by the batteries is used up in lighting the bulb. $N * E_{\mathbf{AA}} = P_{\mathbf{bulb}} * Life$ …(2) Where $P_{\mathbf{bulb}}$ is power rating of the flashlight bulb, and Life is the time that a new set of batteries will take before they die out, and N is the number of batteries in a flashlight.
Parameters and Calculations	N = 2 (number of batteries) $P_{\mathbf{bulb}}$ = 1 Watts Life = 2 hours $E_{\mathbf{AA}} = 1 * 2 * 3600 * 0.5 = 3600$ J $V_{\mathbf{transistor}}/V_{\mathbf{AA}} = 2$ $E_{\mathbf{transistor}}$ = 7200 J

5 Open Issues

In section 3, we mentioned estimation model building and direct parameter estimation as the two key processes underlying BotE reasoning. Our approach is to use similarity to guide both of these processes. Structural similarity, retrieval and generalization form the substrate for this kind of reasoning. What follows is a discussion of important issues in trying to extend these theories to handle quantities. We believe that answers to these questions will form an account of the *sense of the quantitative.*

Our computational account will require extending existing computational models of analogical processes. The structure-mapping engine (SME) [Falkenhainer *et al* 1989] is a computational model of structure-mapping theory. MAC/FAC [Forbus *et al,* 1995] is a model of similarity-based retrieval, that uses a computationally cheap, structure-less filter before doing structural matching. SEQL [Kuehne *et al*, 2000] provides a framework for making generalizations based on exposure to multiple exemplars. With a large number of examples, generalizations will serve to ease the organization of information, and also help in defining typicality and representativeness with respect to parameter values. In order to use experiential knowledge to guide BotE reasoning, SME, MAC/FAC and SEQL have to be extended so that they can make sense of quantitative information. That is, they already can handle representations with numerical parameters, but similarity in aligned numerical parameter values does not affect the perceived similarity of the descriptions compared. Here are some of the issues that are involved:

How do quantitative dimensions factor in our similarity judgments? In our example with the battery, why do we think that an AA battery is more similar to the 9-volt than a car battery, for example? Because we intend to come up with quantitative answers, the similarity comparisons that help us retrieve the relevant examples must take into account the quantitative dimensions in the representations in the first place. Markman and his colleagues have shown in many different experiments that people value aligned differences to be more important for comparison than non-aligned differences. An important question that remains to be explored is in the case of more than one aligned dimension, are all of them equally important, or can one deduce relative importance from structural representations?

What are the quantitative inferences that analogy sanctions? In the direct parameter estimation task, given a base description with a missing value on a dimension, after we retrieve one (or more) matches for which the value on that dimension is known, what kind of strategies do we use to surmise the value for the unknown in our original scenario. This is an interesting question, as it is not necessary that we have an overall match to make estimates along a certain dimension only; and a good match does not mean that all the aligned (numeric) dimensions in the base and the target are equally close.

How do we generalize along quantitative dimensions? In solving the battery example, for example, people say things like "1 Amp is too high a current for a walkman." For domains like the price of a computer, for example, there is no formal way to carve the parameter space into qualitatively distinct regions. Yet, with exposure to multiple examples, we sharpen our notions of what it means for a personal computer to be cheap, medium-range, or expensive. For most of dimensions like the sizes of objects, price of particular consumer goods, etc., we typically encounter multiple different values for a particular parameter whose statistical distribution is unknown to us. To be able to estimate a reasonable value for the parameter in a scenario, one would need to have a notion of what values represent the central tendency, and which are the outliers, and so on. Peterson and Beach (1967) review a number of studies that show that we are equipped with *intuitive statistics* that helps us make such judgments. We are planning to extend SEQL to accumulate distribution information about the parameters assimilated into a generalization.

The primary underlying issue in the above questions is that SME, MAC/FAC and SEQL operate on symbolic, relational representations[3]. We believe a key part of the solution to the above questions lies in figuring out the right

[3] Which means 99 and 100 are as similar/different as 99 and 10000, when treated as symbols, they are both non-identical symbols, but numerically, the differences in magnitude are quite different.

representations for quantity and principles for generating such representations based on experience. As for the former, Qualitative Process Theory [Forbus, 1984] proposed the quantity space representation, where a quantity value was represented by ordinal relationships with *limit points* (points on a scale *where things change*, e.g. Boiling and Freezing Points of a liquid). QP theory showed that such a representation is quite powerful, at the same time allows for expressing incompleteness in our knowledge. There is psychological and linguistic evidence, albeit indirect, that supports the quantity space representation [e.g., Brown and Siegler, 1993]. The notion of limit points might be far more general than dynamical situations. A generalized notion of limit points that extends to examples like cheap/expensive, etc., is what we call *structural limit points*. The idea is that various quantities are relationally tied to each other, or things in world come in structural bundles, and the structural limit points are discontinuities in this structure of relationships.

We are not saying that our internal representations of quantity are purely numeric, or purely symbolic. Numbers are a very powerful representation that can capture as much fine-graininess as one wants, and support operations like the ability to compare quantities and arithmetic across different quantity spaces. A representation of quantity that captures common-sense reasoning will have to support these types of tasks. One of the important things in estimation is the ability to compare quantities. If the parameter itself is not known, then finding a comparable parameter, e.g., one might think of the ceiling as 1.5 times the height of a person, so about 10ft is a reasonable estimate. Guerrin (1995) presents a scheme to map a quality space onto the set of integers so that one can define arithmetic, and with the refinement and abstraction operator, symbols from different quality spaces can be compared. We think an approach like that might be helpful in mapping between qualitative and quantitative scales.

6 Summary

In this paper we have proposed a similarity-based model of back of the envelope reasoning. We propose that the same processes are used in both everyday common sense reasoning and in scientific and engineering reasoning. We also propose that these processes are highly experience-based, using within-domain analogical reasoning and similarity to retrieve, apply, use, and generalize from specific examples and previous problem-solving experience. This model of qualitative reasoning relies heavily on analogical reasoning, and is equipped with a strong sense of quantitative dimensions. We suspect this to be at the heart of common sense reasoning about the physical world.

We are currently exploring this model by using our analogical processing software (SME, MAC/FAC, and SEQL) to create a BotE problem solver. This involves developing a corpus of examples, including descriptions of objects, situations, and behaviors with quantitative parameters. The BotE problem solver we are building will store the solutions it derives in its memory, to model the accumulation of problem-solving expertise.

Acknowledgments

This research is supported by the Computer Science Division of the Office of Naval Research.

References

Berleant, D., and Kuipers, B. 1997. Qualitative and quantitative simulation: bridging the gap. *Artificial Intelligence Journal*, 95(2): 215-255.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100*(3), 511-534.

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

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

Forbus, K. and Gentner, D. (1997). Qualitative mental models: Simulations or memories? *Proceedings of the Eleventh International Workshop on Qualitative Reasoning*, Cortona, Italy.

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

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

Guerrin, F. (1995). Dualistic algebra for qualitative analysis. In *Proceedings of the 9th international workshop on Qualitative Reasoning*, Amsterdam, Holland.

Harte, J. (1988). Consider a spherical cow: A course in environmental problem solving, University Science Books, Sausalito, CA.

Kuehne, S., Forbus, K., Gentner, D. and Quinn, B.(2000) SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of CogSci 2000*, August, 2000.

Linder, B.M. (1991). Understanding estimation and its relation to engineering education, Ph.D. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology.

Markman, A. B., & Gentner, D. (1993c). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431-467.

O'Connor, M.P., and Spotila, J.M. (1992). Consider a spherical lizard: Animals, models and approximations, American Zoologist, **32**, pp 179-193.

Peterson, C.R., and Beach, L.R. (1967). Man as an intuitive statistician, *Psychological Bulletin*, **68**(1), pp 29-46.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function, I. *Psychmetrika, 27*(2), 125-140.

Tversky, A. (1977). Features of Similarity, *Psychological Review* **84**(4), pp 327 - 352.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases, *Science*, **185**, pp 1124-1131.