A Study of Diagrammatic Reasoning from Verbal and Gestural Data

N. Hari Narayanan[1] and **Masaki Suwa** and **Hiroshi Motoda**
Advanced Research Laboratory, Hitachi Ltd.
Hatoyama, Hiki-gun, Saitama Pref. 350-03, Japan.
{narayan,suwa,motoda}@harl.hitachi.co.jp

Abstract

This paper reports on an exploratory study of diagrammatic reasoning. Concurrent think-aloud protocols and gestures of subjects solving a set of device behavior hypothesis problems presented as labeled diagrams were collected. In addition to analyzing verbal protocols, the gestures and marks made by the subjects were examined and used to annotate encoded verbal data. A model of diagrammatic reasoning in this task is proposed and compared with results of analyzing the protocols. Besides lending support to results of previous experimental studies, this study also revealed some interesting aspects of diagrammatic reasoning that merit further investigation.

Introduction

Diagrammatic reasoning may be defined as the act of reasoning, and solving problems, using diagrams as external representations. Though diagrams are static representations, reasoning with them may involve mental simulations of behaviors that change the configuration depicted in the external diagram. Therefore, diagrammatic reasoning may well involve cognitive processes that are imagistic in nature. Whereas mental imagery has a colorful history of research in Cognitive Science, research on diagrammatic reasoning is of more recent vintage (Hegarty, 1992; Hegarty & Sims, in press; Larkin & Simon, 1987; Lindsay, 1988). In this paper we report on an exploratory study of diagrammatic reasoning that was conducted as part of a long-term research program on visual reasoning. The task studied, that of hypothesizing device behaviors from labeled schematic diagrams, is first described. Second, we propose a model of how people might execute this task. Third, the method and results of the protocol analysis study are presented. In the final section we interpret the results within the framework of the model and discuss issues of diagrammatic reasoning that merit further research.

Hypothesizing Behaviors from Diagrams

The diagrammatic reasoning task that we considered in this study is the following: given a schematic diagram of a physical device depicting the spatial configuration of its (labeled) components[2] and an initial behavior, hypothesize the potential behaviors of the device in terms of the behaviors of its components.

This study is part of a larger effort to understand how diagrams are used in reasoning. We are interested in how diagrams support problem solving by directing attention and facilitating visualization, and reduce search during problem solving by indexing relevant knowledge. The task of hypothesizing how a device might behave solely from its diagram is appropriate for this because of its following characteristics:
(1) A device's diagram depicts the spatial layout and configuration of its components. This depiction contains cues that influence how focus of attention[3] is shifted during problem solving.
(2) Behaviors of physical devices include many spatial behaviors such as sliding, tilting, rotating, filling, and deforming. Therefore such problems are ideal vehicles for studying visualization during problem solving.
(3) For problems involving physical devices, only part of the knowledge required to solve them can be gleaned from the diagram. Hence such problems are well suited for studying the interaction of conceptual (non-diagrammatic) knowledge and diagrammatic knowledge during problem solving. The indexing of relevant conceptual and inferential knowledge is a particularly interesting aspect of this interaction.

A Model of Behavior Hypothesis by Diagrammatic Reasoning

The main goal of our experiments was to characterize how visual information from the diagram and conceptual information (prior knowledge) interact, and influence the direction of reasoning, during problem solving. Though solutions that the subjects provided were not always complete and contained inaccuracies, the analyses we carried out – both of the task and the data collected – indicated that the diagram plays two important roles during problem solving.

[1]Current address: Knowledge Systems Laboratory, Stanford University, 701 Welch Rd., Palo Alto, CA 94304. narayan@ksl.stanford.edu

[2]We use the term *components* to mean components, individual parts of components, and substances.

[3]We use the term *focus of attention* to mean an element in the diagram (a component, a part of a component, or a substance) or a conceptual aspect of an element in the diagram (such as the *mass* of a component) that is being heeded at any given moment during problem solving.

- The diagram facilitates the indexing and recall of both conceptual knowledge regarding components and inferential knowledge using which the reasoner can generate new hypotheses.
- The diagram supports a visualization of hypothesized spatial behaviors of components, which in turn enables the reasoner to detect effects of these behaviors.

Based on an analysis of this task and a preliminary examination of subjects' verbal reports, we developed the cognitive process model shown in fig. 1. It explicates a reasoning strategy for solving qualitative behavior hypothesis problems from diagrams. Reasoning starts with a component and its behavior mentioned in the initial condition, and proceeds in cycles. In later cycles, a component and its behavior to focus on will be selected from among the hypotheses in short term memory. The diagram facilitates the indexing and recall of relevant knowledge in two different ways: (i) attending to a component may cue relevant properties of it, information regarding which is either recalled from long term memory or retrieved from the diagram, and (ii) configurational and shape information about components from the diagram together with prior knowledge about components and behaviors allow the indexing and recall of inferential knowledge. New hypotheses are generated in three ways: (i) by deliberating about effects of non-spatial behaviors, (ii) by observing the diagram to locate connected/contacting components and deliberating about how these will be affected by spatial behaviors, or (iii) by mentally visualizing spatial behaviors, detecting interactions among components that result, and deliberating about effects of these interactions. In each of these cases, the application of the recalled inferential knowledge creates new hypotheses in short term memory. For the study reported here, we focussed our analysis on two issues that this model raises: the order in which subjects shift their focus of attention during problem solving and the recall of relevant knowledge cued by spatial information from the diagram.

Method

Subjects. Three adult high school graduates volunteered as subjects.

Materials. The subjects were seated at a table and presented with one sheet (per problem) containing a labeled diagram with an initial condition and instructions written below the diagram. A pen was kept on the table. The subjects were told that they could use it to point or draw on the problem sheet.

Procedure. All subjects attended an initial session in which concurrent think-aloud verbal reporting (Ericsson & Simon, 1983) was explained and illustrated by the experimenter. Each subject attended two problem solving sessions lasting approximately 45 minutes each, separated by a week. Subjects were asked not to discuss the experiments among themselves during this period. In each session a subject was first given a general instruction sheet that explained what was expected of them in terms of think-aloud reporting. These instructions followed the guidelines presented in (Ericsson & Simon, 1983). They were then given three training problems followed by the actual problems. The four actual problems we used are shown in fig. 2. Concurrent verbal and gestural data were collected. Verbal reports (in Japanese) were tape-recorded and gestures with hands and pen were video-taped.

Analyses. The verbal reports were transcribed verbatim, and translated into English. Gestures and drawings that the subjects made were examined using both the video recording and the problem sheets on which subjects drew. Gestures were categorized as *pointing*, *movement-indicating*, or *drawing*. Pointing gestures were further classified as *pointing to a component*, *pointing to an area*, or *tracing a component boundary*. Movement indicating gestures were divided into *moving the tip of the pen*, or *using fingers and/or hands and/or the pen to indicate some spatial behavior of a component*. Movements of the pen's tip were subdivided into three classes: *moving from one location in the diagram to another*, *moving from a component along a direction*, or *moving from one component to another*. Drawing gestures were divided into *drawing the new location and/or orientation of a moving component* or *copying a component*. The verbal protocols were then annotated with encoded gestures. Gestures appeared concurrently with verbalizations or during pauses.

The annotated protocols were segmented. A segment was defined to be a meaningful clause (or clauses) separated from the rest of the transcript by pauses, commas, expressions such as "hmmm", "let me see", etc., question marks, or connectives, or a gesture occurring during a pause. Table 1 gives the number of segments obtained per protocol. Based on its verbal and gestural content, each segment was tagged with one or more of the following labels: diagram observation (DO), visualization (V), recall of known/given information (R), and inference (I). The segments were encoded using four types of expressions: (i) *Focus(components)* lists components that constitute the focus of reasoning in a segment. We assumed that a subject was focusing on a component if it was explicitly mentioned in a segment and/or gestures involving that component (pointing, tracing the boundary, drawing, simulating its motion with the pen, etc.) were made within that segment. We assumed that a subject was focusing on some conceptual aspect of a component if an explicit mention of it appeared in a segment. (ii) *Behavior(component; behavior; type)* describes any given or hypothesized component behaviors mentioned in the segment. (iii) *Knowledge(statements)* contains a restatement of any facts, assumptions, conditions or descriptions of component states that the segment contains. (iv) *Comment(statements)* consists of any elaborations or inferences by the encoder. The encoding vocabulary for the first two types of expressions consists of all component labels, two words for characterizing the type of a behavior – *known* or *hypothesis* – and terms for component behaviors (e.g., move-up, rotate-clockwise, etc.) derived from analyzing the problems and examining the verbal protocols. A limitation of this en-

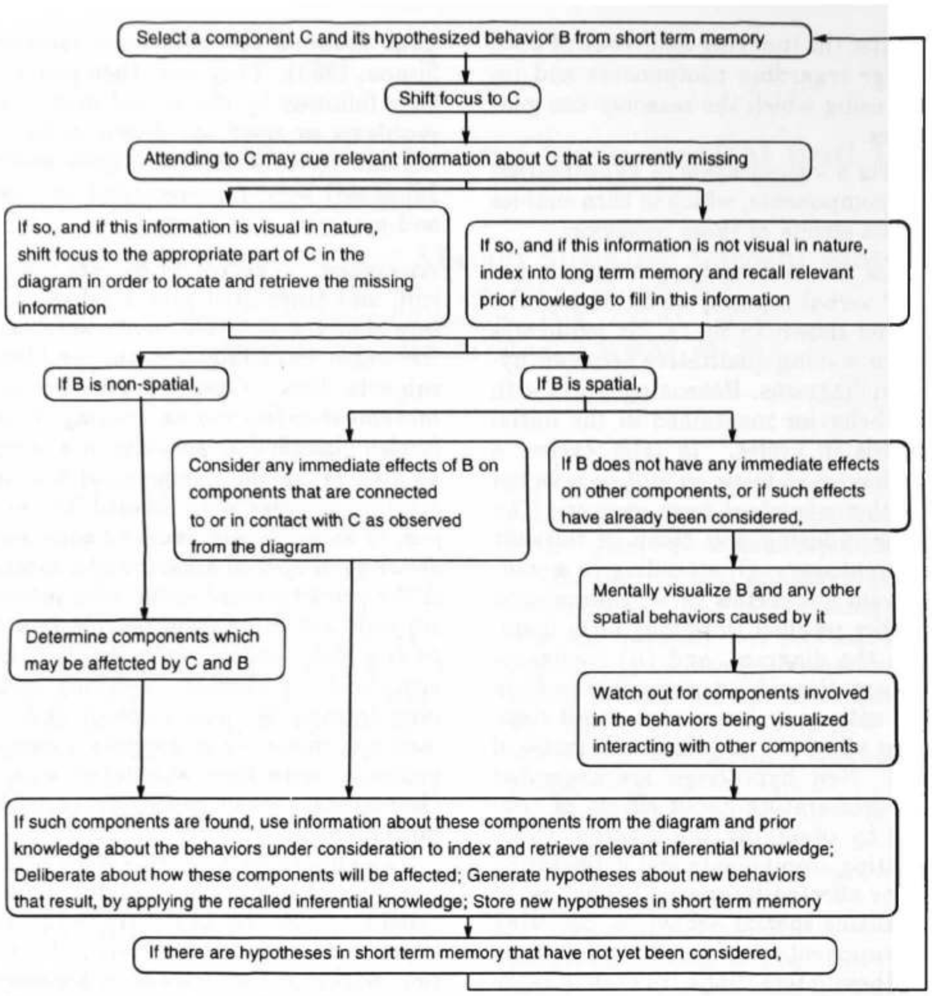


Figure 1: A process model of behavior hypothesis from device diagrams

coding process was that gestures and verbalizations were encoded by one person each. So a measure of the reliability of the encoding, such as inter-coder agreement, is not available.

Results

Our model predicts that reasoning will start with the given initial condition, and will proceed with attention focussed on one or more components. The verbal reports show that reasoning began with the initial condition in all cases.

When attention is focussed on multiple components, the model predicts that these will either be connected/contacting components or components related by the propagation of causality. We counted the number of distinct components indicated by each individual verbalization[4] and accompanying gestures. This number ranged from 0 to 9. In 95% of the verbalizations with more than one component, the components could be classified as connected/contacting components or components causally related by their spatial behaviors. This finding supports the hypothesis in (Hegarty, 1992) that subjects decompose the representation of a device into smaller units corresponding to the machine components during diagrammatic reasoning. Our study suggests that such units may contain more than one component and that this decomposition may be guided by spatial adjacency, causality, or both.

The model predicts that in each cycle the component and behavior selected from short term memory and the hypotheses subsequently generated will be causally related. For a device whose operation can be described in terms of a linear sequence of causally connected events, this means that each pair of successive hypotheses will have a cause-effect relationship. However, for a device whose operation involves branching component behaviors (such as one whose operation can be described by a tree of causally connected events), the model's implication is only that each hypothesis in a sequence generated according to this model can be identified as the effect of some preceding hypothesis in the sequence. The structure and instructions of our problems encouraged subjects to consider many different, possibly branching, be-

[4] A sentence, a question, or a set of clauses separated from the rest of the protocol by pauses at both ends. Such verbalizations were split into multiple segments during the encoding process.

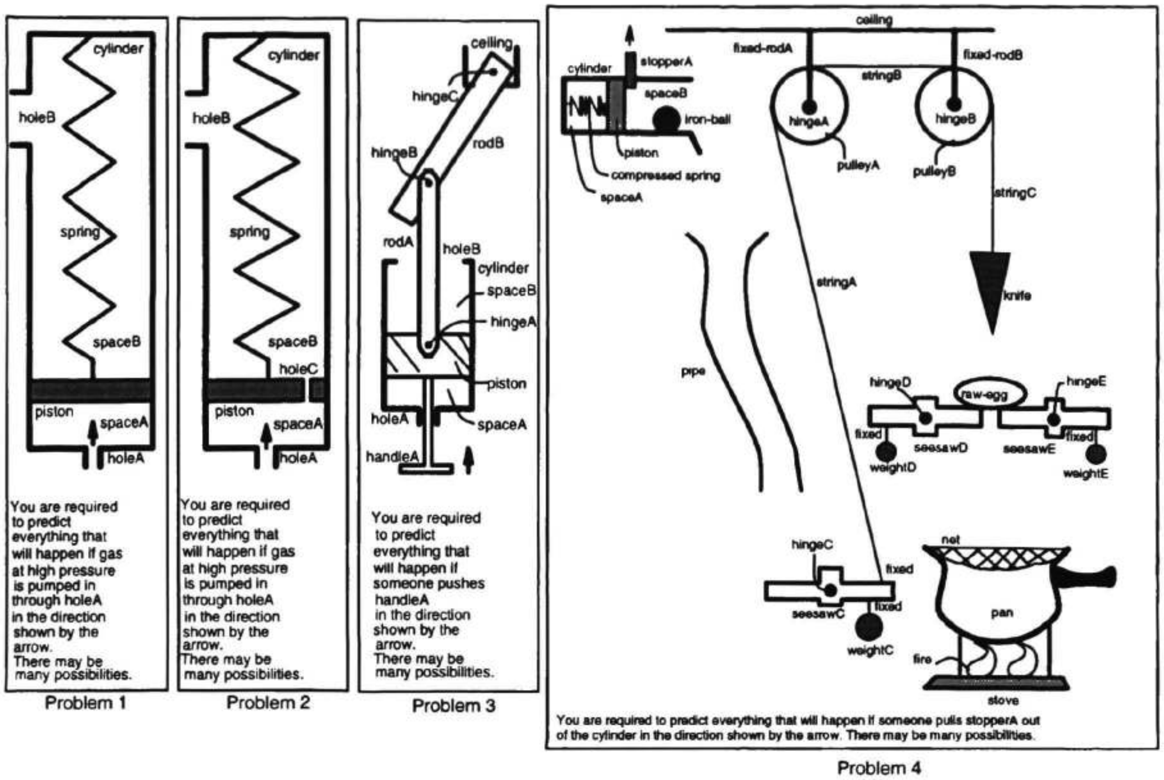


Figure 2: Problems used in this study

haviors. However, their hypotheses appeared in the verbal reports as linear sequences. This means that even if subjects were following the direction of causality in their reasoning, successive hypotheses may not always be related causally. Therefore, we computed the percentage of times a hypothesis about a correct event in the operation of the device, which has another event as its cause and which is appearing in the sequence of hypotheses for the first time, was in fact preceded in the sequence by its cause[5]. This averages to 88% over all 12 protocols. This also supports the conclusion that diagrammatic reasoning about mechanical devices typically proceeds along the direction of causality (Hegarty, 1992).

The model predicts that focus shifts are mediated by three factors – search for information, connectivity/contacts, and visualization. Component connectivity and contacts, detection of spatial interactions among components during visualization of a hypothesized behavior, and search for information were indeed found to be factors, in decreasing order of importance, influencing focus shifts during problem solving. However, these three accounted for only 52% of all focus shifts, indicating that there must be additional factors guiding the shifts of focus.

Focus shifts. Focus shifts were induced from the protocol data based on contents of verbalizations and specific gestures. Therefore, it is a coarser measure than eye fixation data. It was found that connectivity/contacts, visualization, and search for information explain about half of the observed focus shifts. Table 2 summarizes results from analyzing the number and types of focus shifts found in the segments[6], and the different types of focus shifts are explained below.

Connectivity/contacts: A focus shift based on connectivity or contact occurs when a subject, after deliberating about a component and its behavior, starts to heed another component that is connected to or in contact with the former one. Such shifts typically occur when the behavior under consideration affects the shape, position, or orientation of the component. Such shifts of focus guide reasoning along the direction of causality. This appears to be a very common type of focus shift since 39% of all focus shifts in the 12 protocols could be classified into this category.

Visualization: A focus shift based on visualization occurs when a subject, during deliberations about a component and its behavior accompanied by verbalizations or ges-

[5] The formula used was "(X/X+Y) times 100" where X = the number of correct first-time hypotheses with known causes that appeared earlier in the sequence and Y = the number of correct first-time hypotheses with known causes which did not appear earlier in the sequence.

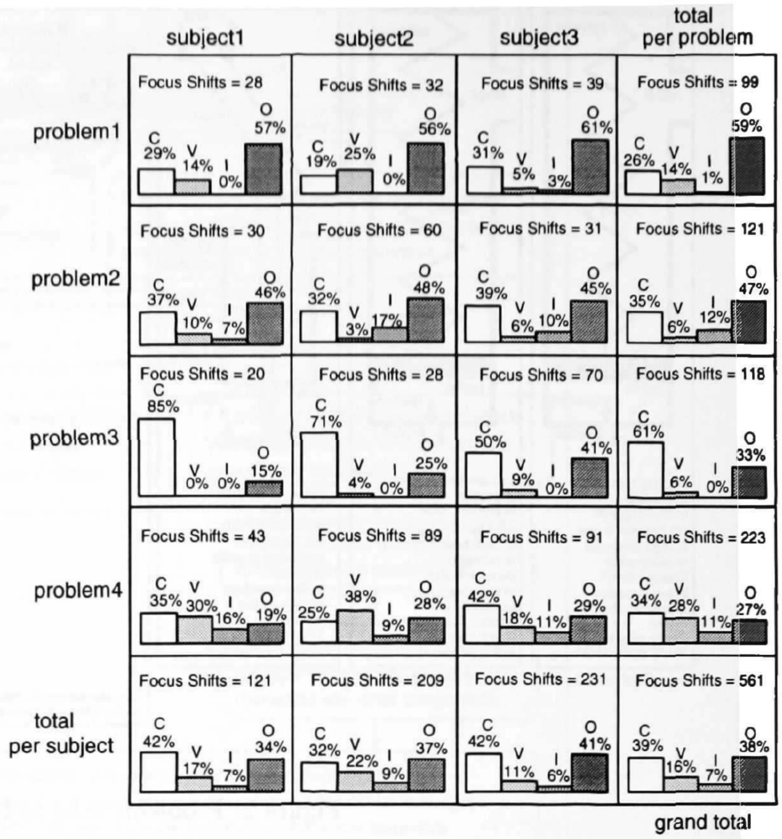
[6] The following abbreviations are used in this table: C - the number of focus shifts explained by component connectivity or contacts; V - the number explained by detection of spatial interactions between components during visualization; I - the number of focus shifts explained by search for information; and O - number of focus shifts that could not be classified as C, V or I.

Table 1: Number of Segments

subjects	1	2	3	total per problem
problem1	26	31	43	100
problem2	25	40	27	92
problem3	28	46	66	140
problem4	36	69	86	191
total per subject	115	186	222	523

grand total

Table 2: Focus Shift Data



tures indicating mental visualization, starts to heed another component that, while not connected to or in contact with the original one, has the potential to interact with the original component as a result of its behavior. Such shifts typically occur when subjects are deliberating about deformation or motion, after considering immediate effects due to connectivity/contact. A shift of focus from the iron ball to the pipe in problem 4 is a typical example. These shifts also guide reasoning along the direction of causality. Among all focus shifts in the 12 protocols, 16% could be classified into this category.
Search for information: Attending to a component may cue prototypical knowledge that a subject has about that component. The indexing and recall of such knowledge (about the component's typical parts or function) may reveal that information about some aspects – visual aspects such as whether a typical part of the component is present in the diagram and conceptual aspects such as the ***purpose*** of a component – of this particular component is currently missing. This will cause a shift of the subject's focus of attention to the corresponding element/location in the diagram (in the case of a visual aspect) or to the corresponding conceptual aspect. Unlike the previous two types of focus shifts, the occurrence of these focus shifts depends on the type of component being considered rather than its behavior. An example is the shift of focus from the left end of seesawC to weightC in problem 4. In this case, considering the left end of a horizontal seesaw might have created an expectation regarding the presence/absence of a balancing weight at the other end, resulting in the focus shift. Since parts of a complex component are typically connected together, it can be difficult to separate this type of focus shifts from those due to connectivity. However, when a shift occurs to a conceptual aspect, or when it bypasses intermediate parts or selects one part when there are other parts that could have been selected based on connectivity alone, we feel that classifying such shifts into this category is appropriate. Only 7% of all focus shifts could be so classified. This may be due to the fact that most of the device components in the four problems were quite simple.

Knowledge indexing. Many instances of the recall of relevant conceptual information triggered by diagram elements (e.g., ***mass*** of the ball and ***purpose*** of the knife in problem 4) appeared in the protocols. We also expected to see instances of diagram elements cueing relevant inferential knowledge. One problem in which this was evident was problem 2. All subjects inferred leakage of gas after noticing holeC in the piston. One subject explicitly considered three possibilities – holeC being very small, normal, or very large – and made correct hypotheses regarding the piston's motion in each case. The other two subjects, perceiving holeC to be very small, ignored its effect and hypothesized the same behavior of the piston (upward motion) as in problem 1. This was the only instance that clearly showed diagrammatic information – the presence and perceived size of holeC – cueing relevant inferential knowledge, although the indexed knowledge and resulting inferences differed among the subjects.

General Discussion

In this section we discuss interesting aspects of diagrammatic reasoning that the verbal and gestural data revealed, and questions meriting further investigations. We close on a speculative note about the potential benefits of studying, and computationally modeling, cognitive processes involved in reasoning with diagrams.

Focus shifts. As table 2 shows, connectivity/contacts, visualization, and search for information explain only 52% of the total number of focus shifts observed over all 12 protocols. Some of the others may be explainable in terms of subjects' goals. One particular pattern appeared many times: after shifting focus to a new component and predicting its behavior, focus shifted back and forth between it and the component perceived to be causing this new behavior. An internal goal to confirm the newly made hypothesis may be the cause of this iterative shifting. Uncovering other determinants of focus shifts and using eye fixation data to track focus shifts more finely are issues deserving further research.

Knowledge indexing. We believe that as in the case of geometry problem solving (Koedinger & Anderson, 1990; McDougal, 1993), diagrams facilitate the indexing of inferential knowledge, thereby reducing search, in the behavior hypothesis task as well. However, the present study did not provide many instances of this. Experiments with problems that are designed to reveal this aspect will be part of our future work.

Visualization. Video-taped gestural data - showing gestures that simulate motions with pen, fingers and hands, and drawings of intermediate and final states of components undergoing motion or deformation - makes a persuasive case for mental visualization of spatial behaviors. Cognitive processes underlying such visualizations need to be investigated further.

Reasoning by visual analogy. In the case of problem 2, *all* subjects recalled their solution to problem 1. Subjects solved these two problems in different sessions separated by a week. Nevertheless, they recalled and tried to adapt and fit the earlier solution to the current problem instead of starting from the beginning. Thus, in this case the diagram helped index the solution from a previously experienced problem solving episode. Such visual analogies (Thagard, Gochfeld, & Hardy, 1992) and their computational modeling using visual cases (Narayanan, 1992) will be part of future research.

Principles and models of problem solving derived from studies of visual and diagrammatic reasoning in real world domains – see (Rogers, 1993) for a compelling example – can form the basis of intelligent visualization and instruction systems that, instead of merely presenting pretty displays, can understand and intelligently manipulate these in order to assist and reduce the cognitive load of their human users. An area where this can make a significant impact is instructional technology. Studies of the role of visual reasoning in scientific thought (Nersessian, 1991), differences in visualization skills between individuals (Hegarty & Sims, in press) and between experts and novices, and common mistakes that novices are prone to, can lead to intelligent tutoring systems that facilitate the development of visualization skills in problem solving among students of science (Dreyfus, 1991) and engineering. This motivates our research on diagrammatic reasoning.

References

Dreyfus, T. (1991). On the status of visual reasoning in mathematics and mathematics education. In F. Furinghetti (Ed.), *Proceedings of the Fifteenth International Conference on the Psychology of Mathematics Education* (Vol. I, pp. 33–48).

Ericsson, K. A. & Simon, H. A. (1983). *Protocol Analysis: Verbal Reports as Data.* Cambridge, MA: MIT Press.

Hegarty, M. (1992). Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1084–1102.

Hegarty, M. & Sims, V. K. (in press). Individual differences in mental animation during mechanical reasoning. *Memory and Cognition.*

Koedinger, K. R. & Anderson, J. R. (1990). Abstract planning and perceptual chunks: elements of expertise in geometry. *Cognitive Science*, 14, 511–550.

Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–100.

Lindsay, R. K. (1988). Images and inference. *Cognition*, 29, 229–250.

McDougal, T. F. (1993). Using case-based reasoning and situated activity to write geometry proofs. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Narayanan, N. H. (1992). Imagery, Diagrams and Reasoning. Doctoral Dissertation. Columbus: Ohio State University, Department of Computer & Information Science.

Nersessian, N. J. (1991). How do scientists think? Capturing the dynamics of conceptual change in science. In R. Giere (Ed.), *Minnesota Studies in the Philosophy of Science 15.* Minneapolis, MN: University of Minnesota Press.

Rogers, E. (1993). Visual Interaction : A Link between Perception and Problem Solving. Doctoral Dissertation. Atlanta: Georgia Institute of Technology, College of Computing.

Thagard, P., Gochfeld, D., & Hardy, S. (1992). Visual analogical mapping. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 522–527). Hillsdale, NJ: Lawrence Erlbaum Associates.