Understanding Illustrations of Physical Laws by Integrating Differences in Visual and Textual Representations

Ronald W. Ferguson and **Kenneth D. Forbus**
Qualitative Reasoning Group
Institute for the Learning Sciences
Northwestern University
1890 Maple Avenue
Evanston, Illinois 60201

Abstract

An important problem in the integration of vision and language is comprehending explanatory diagrams, such as those found in science and engineering textbooks. One class of diagrams, which we call *juxtaposition diagrams*, illustrate a physical principle by comparing two similar situations that vary in a carefully chosen way. This paper describes research in progress on a computational model, JUXTA, which analyzes juxtaposition diagrams. JUXTA performs its analysis by finding the interesting differences in a figure, and then relating those differences to differences stated in the diagram caption. By using the visible differences in the figure as reference points for the qualitative relationship given in the caption, JUXTA is able to intelligently label the relevant parts of the figure. JUXTA also critiques the figure for understandability, warning of differences in the figure which may confuse the reader, and noting visible differences in the figure which are irrelevant and may be removed.

1. Introduction

An important problem in integrating vision and language is understanding diagrams. Diagrams are heavily used in explanatory materials to provide concrete examples that facilitate the understanding of new principles. Understanding a diagram involves figuring out how the idea communicated by the text is embodied in the visual properties of the diagram. In Figure 1, for instance, the relationship between the thickness of a bar and its thermal conductance is illustrated by differences between two similar situations. In these situations, most of the properties on the left and the right are visually the same (and hence we surmise that they are physically the same) except that the bar on the left is thicker, and water is dripping off the ice cube on the left more quickly (as indicated by a greater volume of drops). The caption, while drawing attention to the visible differences between the situations, also confirms that the same causal mechanism (i.e. heat flow) operates in both, and indicates how the visible differences are causally related. We call diagrams such as these *juxtaposition diagrams*. They are commonly used in science and engineering texts. This paper describes work in progress on a computational model, JUXTA[1], for comprehending such diagrams.

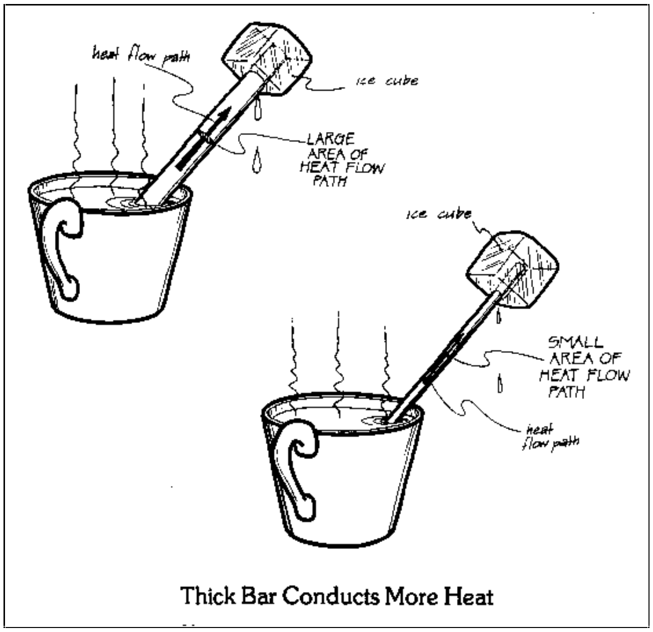
heat flow path
ice cube
LARGE AREA OF HEAT FLOW PATH
ice cube
SMALL AREA OF HEAT FLOW PATH
heat flow path
Thick Bar Conducts More Heat

Figure 1: Figure from *Sun Up to Sun Down* **[Buckley, 1979].**

This research is taking place in the larger context of a project to create a cognitive model of large-scale learning, specifically, a model of the kinds of conceptual change that occur when someone learns from reading a popular science book [Forbus and Gentner, 1991]. Such books typically

[1] JUXTA stands for Juxtaposition Understanding and eXplanation Through Analogy.

To appear at the 1995 AAAI Fall Symposium on Computational Models for Integrating Language and Vision.

describe physical phenomena in qualitative terms, to provide both a working knowledge of an area and to the background needed for more technical training in that area. Such books use diagrams heavily. Our major source text, *Sun Up Sun Down* [Buckley, 1979], contains approximately one diagram per page. Furthermore, many of these diagrams are juxtaposition diagrams (e.g., 17 out of 28 diagrams in three introductory chapters). Thus in modeling the integration of language and vision used to understand juxtaposition diagrams, we are taking an important step toward our larger task of modeling large-scale learning.

A key idea in our account of juxtaposition diagrams is the concept of *alignable differences* [Gentner and Markman, 1994]. An alignable difference is a difference between corresponding parts of two similar situations or entities. For example, in Figure 1 the overall similarity in the left and right situations invites us to place the two bars into correspondence. Because the two bars correspond, the difference in thickness between them is an alignable difference. There is psychological evidence that alignable differences are highly salient, and thus it is natural that they would be exploited in explanatory diagrams.

In juxtaposition diagrams, qualitative laws are illustrated by pairs of alignable differences. The other important visual alignable difference in Figure 1 is that there are larger water drops in the situation on the left. Physically, this means that there must be more water changing phase, which means more heat is flowing in that situation than the one on the right. This difference in heat flow rate is also an alignable difference, albeit not a visual one. Instead, it is stated explicitly in the caption ("more heat"). Understanding this diagram requires noticing and integrating these alignable differences from visual and textual clues into a consistent conceptual account of the phenomena depicted.

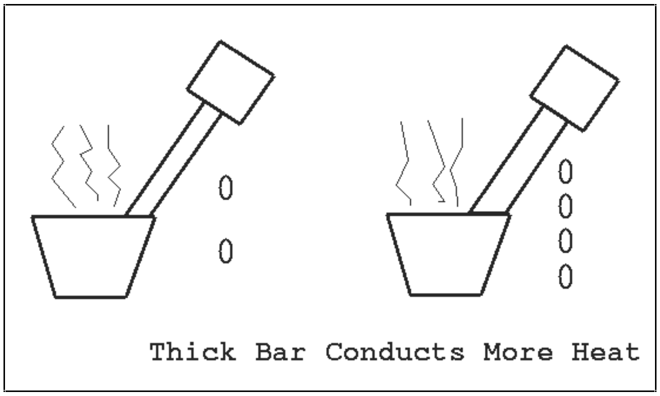

Figure 2: Simplified version of diagram from Buckley, 1979. (*Note:* All figures of this type are direct screen dumps from JUXTA.)

Although our context is modeling "conceptual change by being told", we are using two simpler tasks in evaluating JUXTA as a stand-alone model. First, JUXTA produces labels for the diagrams, to indicate its understanding of the situation. Second, JUXTA also critiques diagram/caption pairs, warning of alignable differences in figures which may confuse readers, and noting visible differences in parts of figures which are irrelevant and can be removed.

The rest of this paper describes how JUXTA works, using its processing of Figure 1 as an extended example.

2. Overview of JUXTA

This section provides an overview of how JUXTA works. JUXTA takes visually simple juxtaposition diagrams as

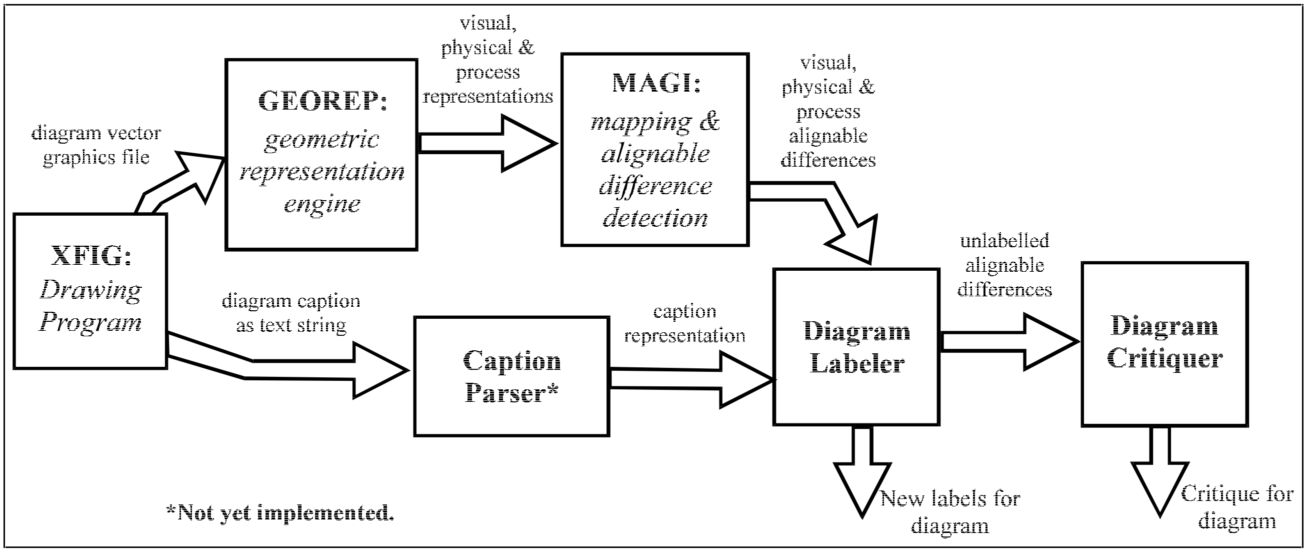
diagram vector graphics file
GEOREP: geometric representation engine
visual, physical & process representations
MAGI: mapping & alignable difference detection
visual, physical & process alignable differences
XFIG: Drawing Program
diagram caption as text string
Caption Parser*
caption representation
Diagram Labeler
unlabelled alignable differences
Diagram Critiquer
New labels for diagram
Critique for diagram
*Not yet implemented.

Figure 3: Architecture of JUXTA

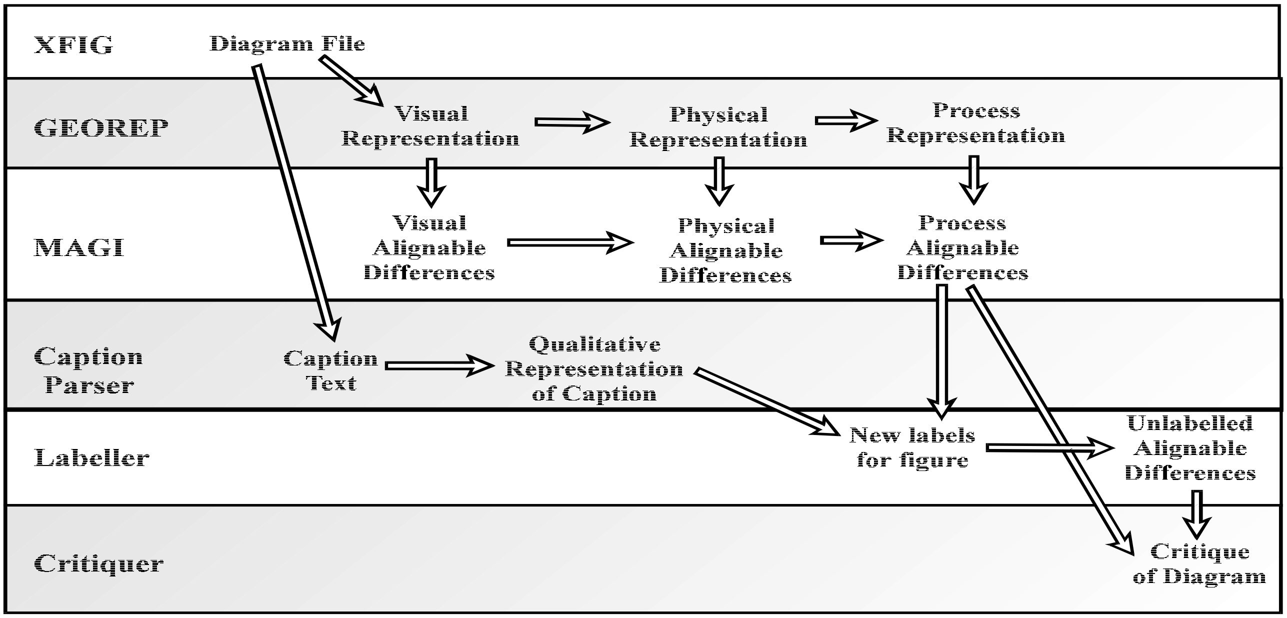

Figure 4: Data flow diagram for JUXTA, arranged in rows by system module

input. (Figure 2 shows the diagram in Figure 1 as seen by the system.) JUXTA provides three kinds of feedback. First, it labels alignable differences in the figure relating to the caption. Second, it critiques differences that interfere with the point of the caption. Finally, it notes when irrelevant alignable differences may be eliminated to increase clarity.

Figure 3 shows JUXTA's architecture. Some of the modules are used with several representations, so the data flow is a bit complicated (see Figure 4).

JUXTA starts with a diagram file, drawn using XFIG.[1] The diagram is read in as a set of geometric objects, including lines, circles, spline curves, and arcs. The diagram file also contains the caption encoded as a text string. Two different processing tracks handle the figure and the caption. The *visual track* (the upper track in Figure 3) processes the geometric elements of the diagram. The *language track* (the lower track, and currently under implementation) parses the caption as a qualitative relationship. The two tracks then meet to handle labeling and critiquing.

Of the two tracks, the visual track does most of the processing in JUXTA. This is because while the language track creates a single qualitative statement from the caption, resolving the references in the caption requires the objects inferred from visual processing. The visual track must detect and represent a number of alignable differences in the figure at multiple levels. First, GeoRep represents the diagram at three different levels--a visual level (e.g. a square), a physical level (an ice cube), and a physical process level (heat flowing into an ice cube) using a set of rules and low-level description routines as described below. At each level of representation, the MAGI symmetry and regularity detector [Ferguson, 1994] maps the representation to itself, and returns the aligned relationships in the diagram (such as two ice cubes or two instances of heat flow). An extension to MAGI detects visual alignable differences between mapped objects (for example, noticing that one metal bar is thicker than another). These alignable differences are integrated by the Diagram Labeler, which creates labels on the diagram corresponding to the two dimensions given in the caption. The unused alignable differences are passed to the Diagram Critiquer, which warns the user if they are potentially confusing.

3. Highlights of how JUXTA works

Here we summarize the critical features of the modules and representations that we believe will enable JUXTA to robustly combine information from vision and language to understand a broad range of diagrams.

3.1. GeoRep: Creating visual and conceptual representations

GeoRep constructs a low-level predicate calculus description of a vector-based graphics file produced by XFIG. The low-level description is based on qualitative, local relationships between proximate shapes. These include different types of line connections, interval relationships between parallel lines, and horizontally or vertically oriented objects. The output of GeoRep can be fed into a variety of systems, to build higher-level visual descriptions based on domain-dependent assumptions about the diagram. In JUXTA a sequence of inference systems

[1] XFIG is a public domain drawing program for X Windows.

Class of object	Visual legend	Salient dimensions (corresponding object dimensions)
Container of liquid	Upright, top-heavy trapezoid	Height and width
Steam or heat	Group of proximate spline curves	Number of curves (amount of heat released)
Metal bar	Oblong, oblique trapezoid with parallel sides	Length and thickness
Ice cube	Square	Width
Water drops	Group of proximate, vertically elongated ellipses	Number of ellipses (amount of water)

Table 1: Visual legends for objects recognized by JUXTA

transform the visual representations of GeoRep's initial processing into conceptual representations of the causal relationships in the situation.

To avoid becoming mired in the problem of visual object recognition, we use a very simple, domain-specific mapping from particular kinds of shapes to types of objects, analogous to legends commonly found in highly schematic diagrams. The particular table we currently use is illustrated in Table 1. While this approach vastly simplifies object recognition, it has the critical feature of retaining the interesting dimensions of the objects in the figure. For instance, a container's height is proportional to the height of the trapezoid that represents it. Relationships between physical objects are recognized through relationships between the representing shapes. For example, immersion of a metal bar in a container of liquid is detected as a shared side between the trapezoid representing the container and the trapezoid representing the metal bar. Because the legends themselves are qualitatively described in terms of shape, they are insensitive to small quantitative changes in placement, and the elements of diagrams are fully compositional. Our approach is only as good as the legend used (for example, JUXTA sees nothing anomalous in the slightly off-center water drops of Figure 2), but it provides plausible scene representations and can be flexibly extended to new objects and relationships.

3.2. MAGI: Using analogy to find aligned differences in a diagram

Once each representation at each level is built, JUXTA must find the interesting differences between the two situations in the figure. For example, in Figure 2, JUXTA should notice that one metal bar is thicker, and that more water drops are falling from the right ice cube. To do this, JUXTA uses MAGI to create an analogical mapping between the maximally similar subparts of the figure, and then compares the mapped objects along salient dimensions. In other words, an analogical mapping constrains the search for differences to those that are based on the aligned parts--thus the term "alignable differences."

MAGI is an extension of SME [Falkenhainer *et al*, 1989; Forbus *et al*, 1994]. SME is a simulation of Structure Mapping Theory [Gentner, 1983], which defines analogy and similarity in terms of sets of correspondences (mappings) between two structured representations. MAGI is based on the insight that symmetry and regularity (visual, conceptual, and mathematical) can be viewed as a special kind of similarity mapping between a description and itself. MAGI's ability to find maximally similar subparts of a figure allows JUXTA to detect the comparison implicit in the figure without being told. Figure 5 shows a MAGI mapping of the figure for the visual level of representation.

JUXTA currently uses a very simple model of alignable differences. For JUXTA, alignable differences are differences along salient dimensions of mapped entities (see Table 1). For example, when two trapezoids are mapped by MAGI, the alignable difference mechanism will then compare those trapezoids to see if they differ in either width or height. When a dimensional difference cannot be

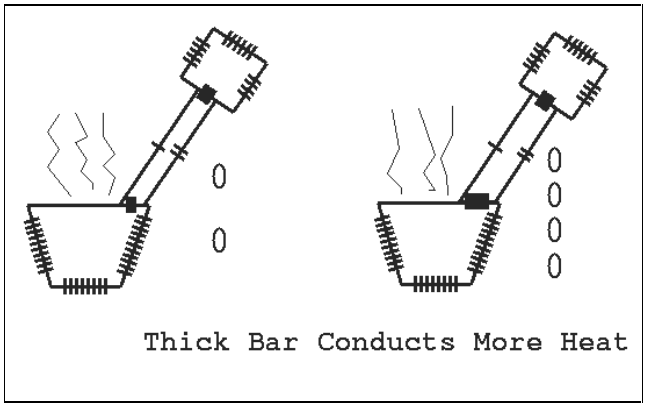

Figure 5: Visual mapping of diagram using MAGI. Mapped parts have an equal number of hash marks.

ascertained directly from the diagram, it may be inferred using qualitative reasoning from dimensional information in the diagram. For example, although the amount of water melting from the ice cube is not directly visible, it can be assumed to be proportional to the number of ellipses in the ellipse group that represents the group of water drops. Using the mapping shown in Figure 5, JUXTA finds the following alignable differences at the visual representation level (Figure 7). While this model of alignable differences is a good starting point, we believe it will need to be broadened to be psychologically realistic.

3.3. The caption representation

Along with the alignable differences returned by the visual track of JUXTA, the language track will build a representation of the key alignable difference/qualitative relationship represented by the caption. Since the parser implementation is still in progress, we currently give JUXTA the representation of the caption directly.

The representation for the example caption “Thick Bar Conducts More Heat,” is shown in Figure 6. The representations use Qualitative Process theory [Forbus, 1984]. It is useful to identify two parts of captions for juxtaposition diagrams, the *antecedent* and *consequent*. In this caption, the antecedent is the difference in thickness of the bars and the consequent is the difference in the rates of heat flow.

JUXTA unifies the caption representation with the physical and process representations of the diagram in order to fill the slots in the caption representation.

```
(metal-bar ?bar1)
(metal-bar ?bar2)
(flow heat ?source1 ?sink1 ?bar1)
(flow heat ?source2 ?sink2 ?bar2)
(qprop (rate
        (flow heat ?source1 ?sink1 ?bar1))
       (thickness ?bar1)) = ?qprop1
(qprop (rate
        (flow heat ?source2 ?sink2 ?bar2))
       (thickness ?bar2)) = ?qprop2
(cause (and ?qprop1 ?qprop2
            (> (thickness ?bar1)
               (thickness ?bar2)))
 (> (rate
     (flow heat ?source1 ?sink1 ?bar1))
    (rate
     (flow heat ?source2 ?sink2 ?bar2))))
```

Figure 6: Representation of caption

3.4. Relating the caption to the visual descriptions

Once the set of alignable differences at each level are detected, and the caption is represented as a qualitative relationship, the diagram labeler attempts to link the aligned differences with the caption representation (Figure 8). This final process is complex, mostly because of the three levels of alignable differences--visual, physical and process-based.

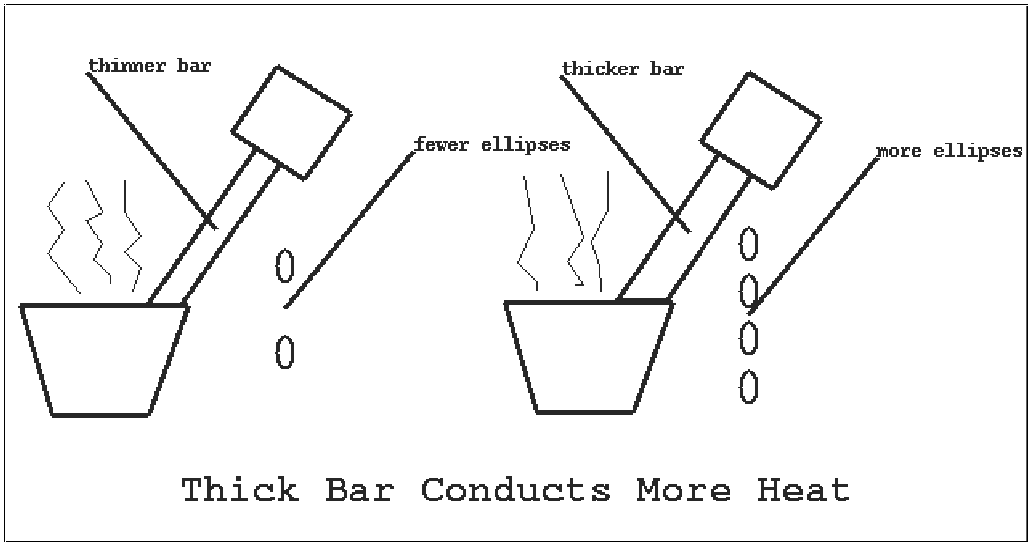

Figure 7: Alignable differences found by JUXTA at the visual level

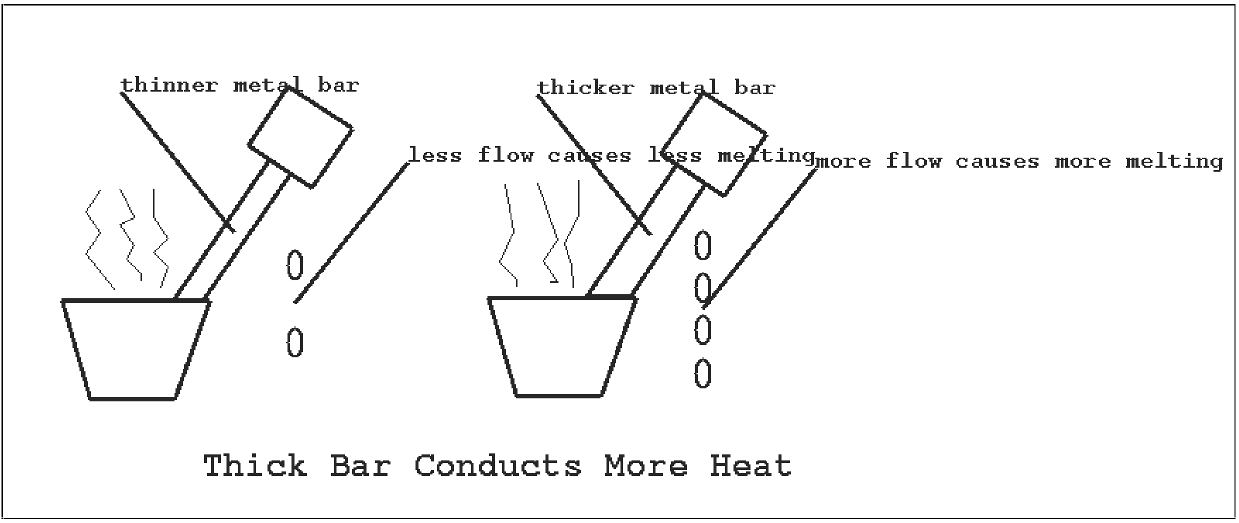
thinner metal bar
thicker metal bar
less flow causes less melting
more flow causes more melting
Thick Bar Conducts More Heat

Figure 8: Labeling for example diagram

Labeling. First, JUXTA attempts to find evidence of the caption antecedent and consequent in the figure, and labels each alignable difference. To label an alignable difference, JUXTA must find a visible referent to point to. In the case when the alignable difference is along a visible dimension (such as the thickness of a bar), the object itself is the referent of the label, and JUXTA points to the shape which represents the physical object. In the case when a caption relationship is not visible (such as heat flow along the metal bar), JUXTA looks for a consequence of the relationship which is visible difference. In the example figure, the difference in heat flow causes a difference in the rate at which the ice cube melts, causing a visible difference in the number of drops (ellipses) , so JUXTA labels this. [1]

Critiquing[2]. After labeling the figure, JUXTA then looks at all remaining alignable differences in the diagram that are not either given in or a consequence of the relationship in the caption. If a remaining alignable difference is not the result of the caption antecedent, but can have an effect on the consequent, JUXTA marks it as potentially confusing. For example, Figure 9 is the same as the example figure, except that the amount of heat rising from the second container is larger than the first container. JUXTA will mark the difference as confusing because the amount of heat from the container implies that the second container may contain a hotter liquid, which would also increase the heat flow rate.

If a remaining alignable difference does not relate to the caption at all, JUXTA will not mark it as confusing, but will note that the alignable difference may not be needed. For example, in Figure 9, JUXTA will note that the middle spline curve in the rightmost group is longer, and making it of equal length may aid diagram interpretation slightly.

4. Conclusion

At present, JUXTA is able to label 3 figures from *Sun Up to Sun Down*, and has been used to parse 3 variants of those figures. With the completion of the diagram critiquer and the extension of the object recognition rules, we expect to be able to parse most of the seventeen juxtaposition diagrams from the introductory chapters of the book, as well as juxtaposition diagrams from other sources.

Currently we are extending JUXTA in three ways. First, implementation is proceeding on a DMAP-style parser to perform language processing. One important change we are making in the parser is the ability to use

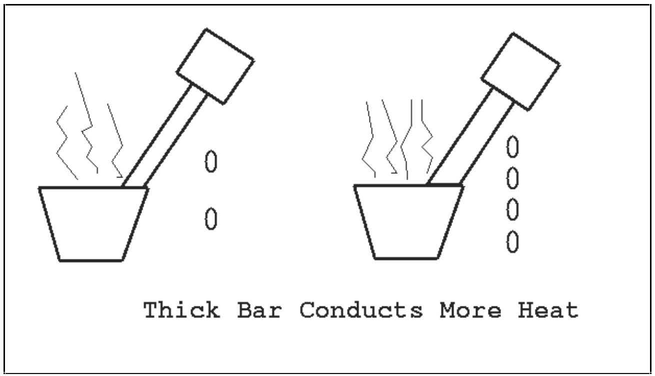

Figure 9

[1] To place the label, JUXTA uses GeoRep’s proximity sensor to find a open location in the figure. It attempts to label aligned differences with labels that are at the same angle and distance, so that the labels themselves also align visually.

[2] The Diagram Critiquer is currently under implementation.

objects identified through visual processing as discourse elements on an equal status with parser-generated representations. Second, we are extending the visual processing in JUXTA to handle a wider range of examples, with the goal of successful operation on all of the juxtaposition diagrams in *Sun Up Sun Down*. Finally, we are looking into ways to make JUXTA generate novel explanatory diagrams based on a given physical situation. In tutoring systems that teach by having the student work through problem-solving tasks in a simulated physical environment, JUXTA may be used to generate juxtaposition diagrams for important physical principles based on the student's current problem solving task, allowing the task to directly motivate the learning of such principles.

Although JUXTA's domain is limited to a particular type of diagram, we believe that many features of its architecture and representations will be applicable to more general problems of understanding diagrams in explanatory material. This is of course an empirical question.

Acknowledgments

Thanks to Laura Allender, John Everett, Edward Lim and Yusuf Pisan for helpful feedback. This research was supported by the Cognitive Science Division of the Office of Naval Research.

References

[Buckley, 1979] Shawn Buckley. *Sun Up to Sun Down.* New York: McGraw Hill Book Company, 1979.

[Falkenhainer *et* al, 1989] Brian Falkenhainer, Kenneth D. Forbus, and Dedre Gentner, D. The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence*, *41,* 1-63, 1989.

[Ferguson, 1994]. Ronald W. Ferguson. MAGI: Analogy-based encoding using regularity and symmetry. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* 1994.

[Forbus, 1984]. Kenneth D. Forbus. Qualitative process theory. *Artificial Intelligence*, *24,* 85-168, 1994.

[Forbus *et al*, 1994]. Kenneth D. Forbus, Ronald W. Ferguson, and Dedre Gentner. Incremental structure mapping. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 1994.

[Forbus and Gentner, 1991]. Kenneth D. Forbus and Dedre Gentner. Similarity-based cognitive architecture. SIGART Bulletin, Vol. 2, No. 4., 66-69, 1991.

[Gentner, 1983] Dedre Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive Science,* 7, 155-170, 1983.

[Gentner and Markman, 1994]. Dedre Gentner and Arthur B. Markman. Structural alignment in comparison: No difference without similarity. *Psychological Science,* 5 (3), 153-158, 1994.