Qualitative Spatial Reasoning for Geometric Analogies

Emmett Tomai   **Kenneth D. Forbus**   **Jeffery Usher**

Qualitative Reasoning Group
Northwestern University
1890 Maple Avenue,
Evanston, IL, 60201 USA
etomai@northwestern.edu

Abstract

In this paper we provide evidence supporting the hypothesis that qualitative spatial reasoning and representations provide a bridge between the perceptual and the conceptual. We demonstrate that qualitative visual structure combined with analogical processing can produce human-like results on a classic geometric analogies task. Importantly, the bulk of the computations are not particular to the task but are general purpose: we sketch the problems using our existing sketch understanding system, sKEA, which computes visual structure that is used by our existing analogical matcher, SME. The results are significant both as evidence in our ongoing research into general-purpose spatial representation, and in illustrating the utility of *second-order analogical matches*.

1. Introduction

One of the mysteries of human cognition is how we make sense of the world around us. We have powerful visual systems, and it appears that part of their job is to compute descriptions of visual structure (cf. [15,19]) which can be used for recognition and understanding. We have argued previously that qualitative spatial reasoning plays an important role in medium and high-level visual processing [8]. Qualitative spatial representations provide a bridge between vision and cognition, since they seem to be computed via visual processes, but take functional constraints into account. We have been exploring this idea by research on sketching. Understanding sketches is a useful approach to understanding qualitative visual structure because starting with digital ink lets us focus on processes of perceptual organization and ignore image processing issues. Previously we have described techniques we have developed for imposing human-like visual structure on sketches and how that structure enables our software to better model human similarity judgments [9]. We now extend that work by applying our system to a set of geometric analogy problems, drawn from the Miller Geometric Analogy Test, designed to test human intelligence; the same set used by Evans in his work on the system ANALOGY [3]. Like Evans, we view these

problems as non-trivial and useful for the exercising of internal descriptions. Our goal is not to improve on the results of ANALOGY, but to validate the ability of our general purpose spatial reasoning system, not specifically designed for this test, to similarly produce human-like analogy judgments. We demonstrate that it is capable of doing so and that its performance is directly dependent on the elements of qualitative visual structure as well as the availability of common-sense knowledge relevant to the task.

We start by reviewing our approach to sketching and the *sketching Knowledge Entry Associate* (sKEA) [10], an open-domain sketching system used in these experiments. We describe how sketches are represented and the visual structure that we compute over them. Next, we describe the geometric analogy problems we are working on and our strategy for solving the problems via structure mapping. We walk through an analysis of one example in detail to illustrate the processing, and then summarize what we found from the rest of Evans' problems. Finally, we discuss plans for future work.

2. Overview of nuSketch and sKEA

Sketching is a form of multimodal interaction, where participants use a combination of interactive drawing and language to provide high-bandwidth communication. Sketching is especially effective in tasks that involve space, e.g., physical structures or maps. While today's software is far being as fluent as sketching with a person, research on multimodal interfaces has produced interfaces that are significantly more natural than standard mice/menu systems (cf. [2]).

sKEA is designed to enable knowledge entry via sketching. Unlike most sketching systems, which are limited to a narrowly constrained domain, sKEA is open-ended: Any concept in its large knowledge base can be included in a sketch. Specifically, we use a subset of Cycorp's Cyc knowledge base contents, with extensions developed by our group for qualitative and analogical reasoning.

The typical approach in multimodal interfaces is (a) to provide a more natural interface to a legacy software system and (b) to focus on recognition [1,2]. While this approach has led to useful systems, it has

some serious limitations. First, today's statistical recognizers are not very good. Second, speech recognition requires that the vocabulary and grammar can be fixed in advance. This can be reasonable for sketching systems operating in tightly constrained domains, but for sKEA, which is designed to be general-purpose, such *a priori* restrictions are not possible. Third, even if recognition improves to human-level, there is still the problem of providing software with a visual and conceptual understanding of what is being sketched. Such knowledge is crucial for creating knowledge capture and performance support systems.

Our approach in the *nuSketch architecture* [8] is quite different and complements traditional multimodal research. We avoid recognition issues by using clever interface design. We focus instead on providing richer visual and conceptual understanding of what is sketched. sKEA's interface provides ways for users to specify conceptual information about the entities being sketched [10]. sKEA also uses the knowledge base to draw additional inferences about the conceptual relationships depicted in the sketch. sKEA is still a research system, although we have carried out internal experiments where graduate students from other groups were able to use it successfully.

3. Representing glyphs and sketches

This section describes the underlying ontology of sketches that we use. The basic unit in a sketch is a *glyph*. Every glyph has *ink* and its *content*. The ink consists of one or more poly-lines, representing what the user drew when specifying that glyph. The content is a conceptual entity, the kind of thing that the glyph is representing. For example, when a user draws a square, there is an entity created to represent the glyph itself and an entity to represent the square.

While some basic spatial properties of glyphs are computed (described below), we do not perform any detailed shape reasoning on the ink comprising a glyph, nor do we attempt to visually decompose it. We call this *blob semantics* because it focuses on spatial relationships between glyphs rather than detailed reasoning about the visual structure of glyphs themselves. While insufficient for recognition based on detailed visual similarity of specific features, it is an excellent approximation for many kinds of spatial reasoning, whenever the focus is on configural relationships between glyphs.

4. Spatial processing of glyphs

When a glyph is added, moved, or resized, sKEA computes a set of spatial attributes and relationships using two threaded processors. These processors, and the intermediate structures they create and use, are described in detail in [9]. We now discuss the spatial

attributes and relationships that make up the visual structure of the sketch.

4.1 Grouping

sKEA computes two types of automatic groupings: *contained glyph groups* and *connected glyph groups*. A contained group consists of a single container glyph and the set of glyphs that are fully contained within it, possibly tangentially so. The contained group does not include glyphs that are contained within other glyphs in the group. A connected glyph group consists of a set of glyphs that overlap ink strokes with one another. Articulation points can be computed over connected glyph groups, and tangentially connected pairs of glyphs can be noted as such. The algorithms used for computing glyph groups are detailed in [9].

4.2 Positional relationships

Positional relationships are computed pair-wise and expressed in a viewer-oriented coordinate system of left/right and above/below. They are not computed between all pairs of glyphs but rather in local neighborhoods based on adjacency. Positional relationships are not computed between glyphs on different layers. The algorithms used for computing positional relationships are detailed in [9], except for one important update: positional relationships can now be computed between intersecting glyphs, as long as one does not entirely contain the other.

4.3 Size

The computation of glyph size in sKEA assigns each glyph a qualitative size value from the set of *tiny*, *small*, *medium*, *large* and *huge*. Sizes are based on the area of a glyph's axis-aligned bounding box, a coarse but empirically useful approximation. Glyph areas are normalized with respect to either the area of the bounding box around all glyphs on all layers, or the area of the user's view port, whichever is larger. The normalized areas are then clustered into qualitative size values based on a logarithmic scale of the square root of the area. Informal experimentation suggests that this is a reasonable method for the varieties of sketches we have examined thus far.

4.4 Orientation

The orientation of a glyph is represented using the qualitative vector notation of [17]. That is, the sign of the X and Y coordinates are used as the orientation.

5. The Miller Analogies Test problems

The geometric problems we are working are from the Miller Analogies Test, a standardized test of human

intelligence. Using our sketching system we have drawn the problems as they appear in Evans' work.

Each of the problems presents the analogy "A is to B as C is to ?" with five possible answers, numbered 1 through 5. The correct answer is the one that best completes the analogy. These problems are depicted as eight figures, labeled A, B, C and 1 – 5. Each figure is a configuration of shapes, lines and other symbols. Figure 1 shows one example of a sketch made from a diagram in Evans' work. The red labels in the illustration have been added to improve readability.
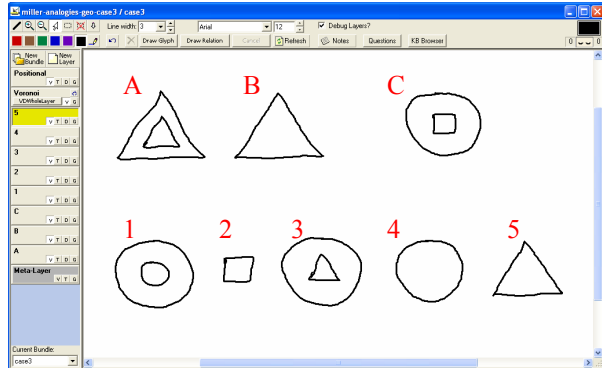
Figure 1. sketch of Problem 3

There are three capabilities a system must have in order to solve these problems: a mechanism for representing the problems to the system, the ability to evaluate the spatial structure of the figures and the ability to evaluate the appropriateness of the analogies proposed by each answer.

6. Representing MAT problems with sKEA

sKEA provides a natural means of entering the geometry problems. We use the layer facility to create eight layers named A, B, C and 1 – 5 which will contain the glyphs that make up each respective figure. Each shape, line and symbol in each figure is drawn as a separate glyph so object segmentation is not an issue.

Since we are not exploring visual recognition, we use the ability to conceptually label glyphs to identify shapes to the system. For example, in Figure 1, the glyphs would be labeled as triangles, circles and squares. This is equivalent to part of the functionality of part I of Evans' ANALOGY system, which recognized lines and the shapes that they make up. It is a reasonable approximation because our concern, as described above, is with understanding the spatial and conceptual relationships between the shapes and lines. However, there are still outstanding issues with object decomposition that are discussed following the experimental results.

7. Solving geometric analogies

As in our previous work, we rely on the *Structure-Mapping Engine* (SME) [4], an implementation of Gentner's structure mapping theory [12] to provide human-like analogical processing. Because SME is domain independent, we are able to focus our investigation on the representation of the problems.

To solve the geometric analogy problems, we use a two-stage structure mapping process, depicted in Figure 2. The first stage is the computation of mappings from figure A to figure B and from figure C to each of the answer figures 1 – 5. This generates six mappings (the example mapping AB and the potential answer mappings C1 – C5) that represent the similarities and differences between their respective pairs of figures. The second stage takes those mappings as input and computes the prescribed analogy from AB to each of the answer mappings C1 – C5. The strongest result from the second stage indicates the correct answer. The second stage is an example of what we call *second-order analogical matching*.
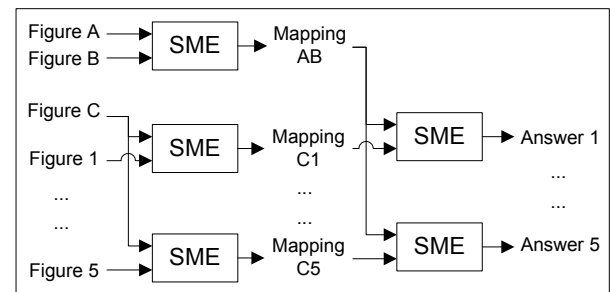

Figure 2. two-stage structure mapping

7.1 First stage structure mapping

SME takes as input the representations of a pair of figures and returns a mapping between them. This mapping consists of *entity correspondences* that match objects, glyphs and layers from the base with corresponding entities in the target, *expression correspondences* that form the multi-level structure of support for the entity matches and *candidate inferences* that project unmatched relationships and features from the base into the target [4]. For a given pair of figures in these analogy problems, the entity and expression matches represent the structure of similarity while the candidate inferences represent *alignable differences*, differences that are connected to common structure [13].

Standard SME operation computes alignable differences only from the base to the target. For our purposes, it is just as important to detect novel relationships and attributes in the target that are not present in the base. We therefore added the ability to compute candidate inferences in the reverse direction

using the same algorithm used to generate standard candidate inferences but with swapped arguments.

7.2 Similarity vs. difference

The first stage of comparison works through the similarities between pairs of figures. As research in analogical reasoning has shown [13], descriptions of differences arise out of comparisons. Because the alignable differences computed as forward and reverse candidate inferences by the first stage are already grounded in the similarities, it is reasonable to suggest that those differences provide all the necessary information for this task. In our experiments with the twenty problems from Evans' work we have passed only the alignable differences to the second stage, omitting the entity and expression correspondences; results so far have shown this to be sufficient.

7.3 Domain knowledge

In order to make meaningful comparisons between conceptual terms such as "circle", "medium sized" and "vertically oriented", the system requires common-sense domain knowledge about those terms. It must know that circles, squares and triangles are all types of shape and have the same kind of knowledge about sizes and orientations. This taxonomic information is contained in our knowledge base as Cyc-style *genls* and *isa* relationships (e.g. (genls Circle GeometricallyDescribableThing)). In comparing sizes there is additional ordering information from smallest to largest, and in comparing orientations there is the concept of rotation from one orientation to the next. These facts form the domain of knowledge necessary for the solution of these geometric analogy problems. We make the knowledge available to the system in the form of general knowledge within our knowledge base.

The system elaborates the results of each first-stage mapping by querying the knowledge base, retrieving knowledge based on the attributes in the mapping and what relationships hold between them. These elaborated descriptions become the input for the second stage.

7.4 Scoring possible answers

The scores used to evaluate possible answers are computed by combining two factors. The first factor is the structural evaluation score computed by SME for the second stage mappings. This factor indicates how structurally similar the input representations are. For these problems, it indicates how strongly the differences in the example mapping are reflected in the answer mapping. The second factor is required because the structural evaluation score does not penalize the answer mapping for having additional differences, *leftovers*, that are not present in the example mapping.

For example, in the sketch of Problem 3 (illustrated in Figure 1) answer 4 could be seen as the removal of a glyph while answer 5 would be seen as the removal combined with a shape change. Clearly the example pair AB shows only the removal. In spite of this, these two answers receive the same structural evaluation score since they both reflect the removal. The shape change is a leftover and should be penalized.

After the second stage mapping, the leftovers are grouped into three types and the counts of each type are multiplied by penalty coefficients and added together. This number is multiplied by a fixed weight and subtracted from the structural evaluation score.

1. A coefficient of 0.5 is applied to leftover attribute expressions. These indicate that an entity in the answer mapping shows a spurious feature difference (shape, orientation or size) when compared to the example mapping. The coefficient is 0.5 because attribute changes always generate a pair of expressions, a candidate inference and a reverse candidate inference, indicating the two different attribute values in question.

2. A coefficient of 1.25 is applied to leftover expressions involving skolems. These indicate an unmatched glyph addition or removal in the answer mapping. Skolems are more heavily penalized because they are less likely to be a false positive (i.e. a result of the limitations of blob semantics or drawing skill).

3. A coefficient of 1.0 is applied to leftover relationship expressions. These indicate that a relationship between entities in the answer mapping appears or disappears when no such change occurs in the example mapping.

8. A detailed example

To illustrate the system's operation and the issues raised by it, we walk through one of the problems, Problem 3, depicted in Figure 1.

In this problem the correct answer is figure 4. The difference between figure A and figure B is the lack of the smaller, inner triangle. Figure C likewise has a contained small square that is lacking in figure 4.

8.1 Sketching the problem

Our first step is to draw the sketch in sKEA. Each shape is drawn as an individual glyph in the proper layer and identified to the system as a triangle, square or circle using those concepts from the knowledge base.

sKEA's visual processing computes size and orientation for each glyph. The larger glyphs are all medium while the smaller are small. Orientation does not come into play since none of the shapes here have distinct major axes. Contained glyph groups are asserted for figures A, C, 1 and 3; no connected glyph groups are found. There are no adjacent glyphs within any of the layers and thus no positional relationships.

8.2 First stage structure mapping

The first stage structural mapping between figures A and B maps together the two larger triangles on the strength of their size and shape and generates a candidate inference proposing that the triangle in figure B should have another glyph inside of it. No reverse candidate inferences are formed.

The first stage mappings from figure C to each of the five answer figures return notably similar results, showing changes in shape and removal of the inner and outer glyphs, as one would expect looking at the problem.

8.3 Second stage results

The second stage mappings correctly identify figure 4 as the answer. Answer mappings from figures 1 and 3 generate candidate inferences and reverse candidate inferences indicating difference in the shape of the inner glyphs. These fail to map with anything in the example mapping resulting in null scores for both. The answer mapping for figure 2 generates a candidate inference indicating the lack of the outer glyph. This fails to map with the lack of an inner glyph, again resulting in a null score. Answer mappings for figures 4 and 5 receive identical structural evaluation scores for reflecting the removal of the inner glyph. However, figure 5 is penalized for having a leftover, the difference in the shape of the outer glyph, and figure 4 is selected as the answer.

9. Experimental results

Due to space concerns we cannot include the sketches of all twenty problems. We present instead a table of results followed by analysis of the findings.

| Correct	1,2,3,4,5,6,7,8,9,11,14,16,18,19,20
Incorrect	10,12,15,17
Ambiguous	13

Figure 3. table of results
An answer is tabulated as correct when our system agrees with the correct answer on the original test, as reported by Evans [3]. Ambiguous answers indicate a tie in scoring.

9.1 Analysis

Problem 10 is the only one of the twenty in which ANALOGY selected an answer different from that selected by the human test makers. Problems 12-20 were run only through part 2 of ANALOGY due to limitations in recognition. The inability of our system to properly solve some of the problems can be traced to four shortcomings, some of which are also visible in problems that were answered correctly. They are the

inability to do axial symmetry, the inability to decompose glyphs (the blob semantics assumption), a lack of hierarchical awareness in positional relationships and the inability to reinterpret the example pair and try a different avenue of attack. We discuss each in turn.

9.2 Axial Symmetry
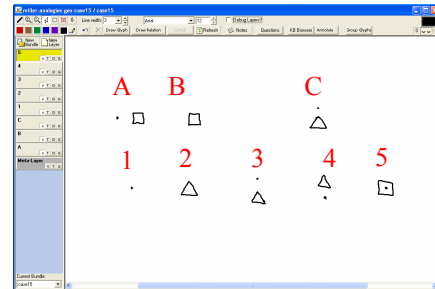
Problem 15

Problem 15, illustrated above, fails because two different positional relationships cannot be analogous. That is, a dot missing from the left does not map to a dot missing from above. This clearly indicates that our positional relationships are too rigid in this respect and must be allowed to compare within frames of reference. Problems 1 and 7, below, also have symmetry issues where the disappearing dots are not properly recognized. In those cases, however, there are sufficient other factors for the system to answer correctly.
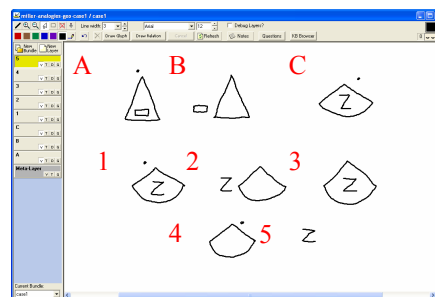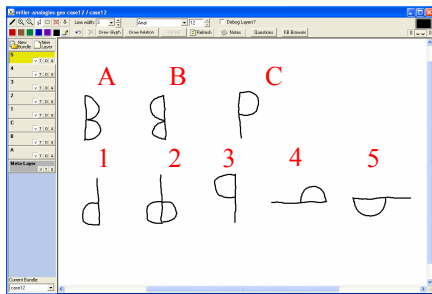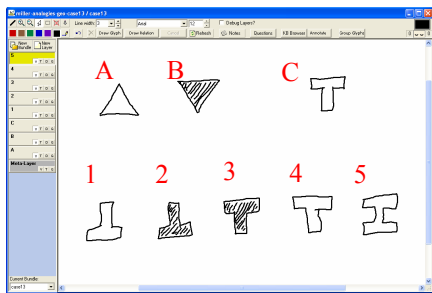

Problem 1
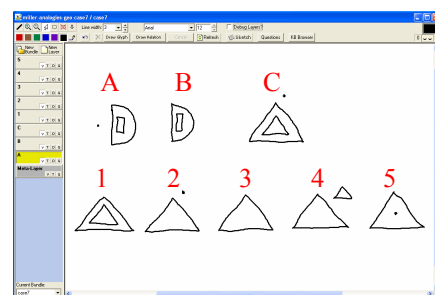
Problem 7

9.3 Glyph Decomposition

Problem 12

Problem 13

We found that comparing glyphs often requires decomposition for meaningful representation. Under the blob semantics assumption, the "letter B" in Problem 12 appears as a thin, vertical glyph. Although the system knows that both figures A and B are instances of the "letter B", it cannot tell that one is mirrored (or rotated) from the other. It therefore sees the example pair as showing no differences at all. In Problem 13, the example is interpreted as shading being added to a triangle; the change in orientation is invisible under blob semantics. Based on this, it can narrow the options down to answers 2 and 3, both of which show shading added, but can go no further and results in a tie.

Many of the Problem cases reflect this issue, although only those two have nothing else to go on. In Problem 18, for example, only answer 3 fits the requirements of having unchanged conceptual shape and orientation. In Problem 2, each line is drawn as a separate glyph which allows the system to see the changes in orientation even under blob semantics.
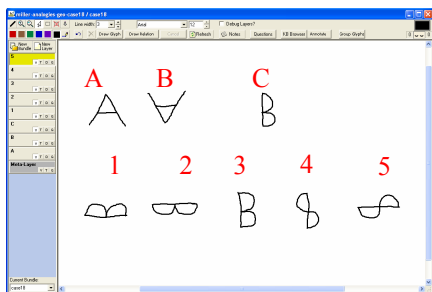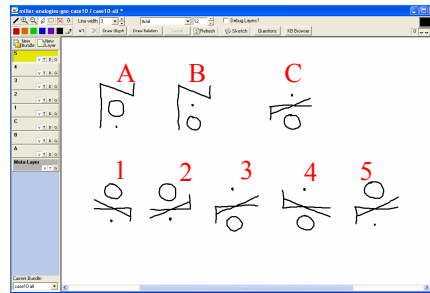

Problem 18
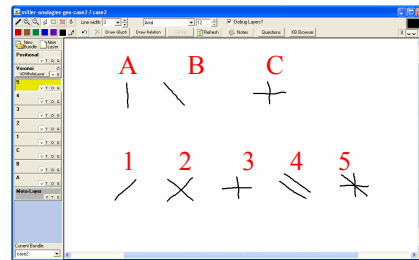
Problem 2

## 9.4 Positional Relationships and Hierarchy

Problem 10

In Problem 10, the mappings between figure C and the answer figures are characterized as differences in shape rather than positioning. Structure mapping gives more weight to relationships than to features and so, in lieu of other evidence, favors maintaining positions. The example mapping on the other hand is characterized as the circle and the dot changing positions. This occurs because positional relationship are localized and the presence of the lightning shaped polyline in figures A and B causes the neighborhood of both circle and dot to look the same (something above, something to the left). When the relationships become ubiquitous like this, they appropriately wash out, leaving other factors, shape in this instance, to drive the result. Because the example pair is characterized as difference in position and the answer pairs as difference in shape, the system rejects all the answers and fails.

This issue suggests that while the neighborhood view of positional relationships is highly effective in mapping the elements of common configurations, it lacks the ability to identify those configurations. Ongoing work with hierarchical grouping of glyphs offers the opportunity to give those relationships stronger structural context so that the proper neighborhoods find each other.

9.5 Reinterpretation
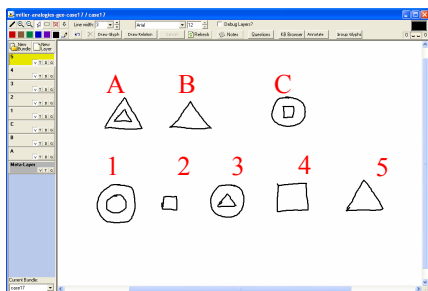
Problem 17

Problem 17 is a recasting of Problem 3, with the original answer removed and replaced with a slightly more difficult one. Instead of an interior glyph missing, the exterior glyph is missing and the interior glyph is scaled up to give answer 4. The example pair supports both interpretations, but the original is clearly stronger. Solving Problem 17 requires backing off the original interpretation and trying another route, something that our system does not do. ANALOGY was able to solve this problem because it does a search of the space and finds the best possible transformation sequence. In our system, we could do an analogous process by running the first stage SME computation multiple times to get every possible mapping. Such an exhaustive search does not however constitute an interesting extension to the system. What it does suggest is the application of different solving strategies and the very difficult question of evaluating the quality of the best available answer without doing full search.

10. Other related work

Evans' classic work was the first to illustrate that machines could do analogy. To fit his program into the punch-card machine available at the time, the geometric processing was done as a separate module, taking coordinates as input and producing symbolic descriptions. Due to limitations in this part of the program, some of the examples reported in [3] actually use hand-coded inputs instead. Subsequent attempts to build on Evans' work that we are aware of all use hand-generated symbolic inputs as starting points (e.g., [18]). By contrast, our model exploits sKEA's built-in qualitative visual structure computing abilities to generate representations from ink input, capabilities which are part of a general-purpose architecture for sketch understanding. sKEA's visual processing evolved from Ferguson's work on GeoRep [5], which operates in an off-line mode on line drawings.

Another significant difference is that Evans construed the problem as creating explicit transformation rules between pairs of figures, which led to ambiguities due to the need to consider alternate possible rules in some cases. Our model illustrates that computing explicit rules is unnecessary: Comparing the similarities and differences is sufficient to explain behavior on the task.

Tight interleaving of the construction of representations with matching is a hallmark of systems from Hofstader's group, including Mitchell's Copycat program [15] and French's TableTop [11]. Unfortunately, each of these systems only operates in the single domain it was designed for, letter-strings for Copycat and table settings for TableTop. The kinds of comparisons that can be made are hand-wired into the system (the Slipnet). By contrast, SME has been used in a wide variety of domains, and automatically figures out what kinds of things can be matched [7].

11. Discussion and future work

We have shown that qualitative representations are a significant element of doing geometric analogies of a kind commonly used in intelligence testing. We have also shown that the set of representations we are working provide a reasonable subset of the representations necessary for solving said problems. Finally, we showed that a two-level analogical processing scheme could capture the phenomena without introducing transformation rules, as Evans did.

Future work will include continued research on visual structure as well as conceptual relationships. We plan to extend our visual processing and experiment with Ferguson's MAGI model of symmetry [5]. We also intend to introduce conceptual grouping as both context for spatial qualities and as a foundation for richer conceptual relationships between sketched entities. Finally, we plan to stretch the boundaries of blob semantics by exploring automatic recognition of known shapes and techniques for the decomposition of blobs into visually meaningful pieces of ink. Ongoing work in this area has proposed a mixture of interactive and automated techniques (cf. [14,20]).

12. References

1. Alvarado, Christine and Davis, Randall (2001). Resolving ambiguities to create a natural sketch based interface. *Proceedings of IJCAI-2001*.
2. Cohen, P. R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. (1997). QuickSet: Multimodal interaction for

distributed applications, *Proceedings of the Fifth Annual International Multimodal Conference (Multimedia '97)*, Seattle, WA, November 1997), ACM Press, pp 31-40.

3. Evans, T. (1968) A Program for the Solution of a Class of Geometric-Analogy Intelligence-Test Questions, *Semantic Information Processing*, 1968, MIT Press.

4. Falkenhainer, B., Forbus, K., Gentner, D. (1989) The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence*, 41, pp 1-63.

5. Ferguson, R. W. (1994). MAGI: Analogy-based encoding using symmetry and regularity, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*.

6. Ferguson, R. and Forbus, K. (2000). GeoRep: A flexible tool for spatial representation of line drawings. *Proceedings of AAAI-2000*. Austin, Texas.

7. Forbus, K., Gentner, D., Markman, A. and Ferguson, R. (1997). Analogy just looks like high-level perception: Why a domain-general approach to analogical mapping is right. *Journal of Experimental and Theoretical Artificial Intelligence* (JETI), 4, 185-211.

8. Forbus, K., Ferguson, R. and Usher, J. (2001). Towards a computational model of sketching. *IUI'01*, January 14-17, 2001, Santa Fe, New Mexico

9. Forbus, K., Tomai, E., and Usher, J. (2003). Qualitative Spatial Reasoning for Visual Grouping in Sketches. *Proceedings of the 17th International Workshop on Qualitative Reasoning*, Brasilia, Brazil, August 2003.

10. Forbus, K. and Usher, J. (2002). Sketching for knowledge capture: A progress report. *IUI'02*, January 13- 16, 2002, San Francisco, California.

11. French, R. (1995). *The subtlety of similarity*. Cambridge, MA: The MIT Press.

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

13. Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45-56.

14. Mahoney, JV, & Fromherz, MPJ (2002). Three main concerns in sketch recognition and an approach to addressing them. *AAAI Spring Symposium on Sketch Understanding*, pp 105—112, March 25-27 2002.

15. Marr, D. (1983). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman

16. Mitchell, M. (1993). *Analogy-making as perception: A computer model*. Cambridge, MA: The MIT Press.

17. Nielsen, P. (1988). A qualitative approach to rigid body mechanics. *Technical Report No. UIUCDCS-R-88-1469*, Department of Computer Science, University of Illinois at Urbana-Champaign.

18. O'Hara, S. and Indurkhya, B. (1995). Adaptation and redescription in the context of geometric proportional analogies. *AAAI-95 Fall Symposium on Adaptation of Knowledge for Reuse*, November, pp 80-86.

19. Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. MIT Press.

20. Saund, E., and Moran, T, (1995). Perceptual Organization in an Interactive Sketch Editing Application. *ICCV '95*.