A Computational Model of the Visual Oddity Task

Andrew Lovett (andrew-lovett@northwestern.edu)
Kate Lockwood (kate@cs.northwestern.edu)
Kenneth Forbus (forbus@northwestern.edu)
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Rd, Evanston, IL 60208-3118 USA

Abstract

Understanding how high-level visual properties are computed is a central problem in perception. Oddity tasks, where participants must identify a stimulus that is distinct in some way from others in an array, provide a method for determining what features are being computed. We describe a computational model of oddity detection that models data by Dehaene et al. (2006) on perception of simple geometric shapes. It starts with virtually the same input stimuli as given to human subjects, and automatically constructs representations. Oddity detection is accomplished by analogical processing, using SME and SEQL. The simulation is able to perform the task, and moreover, provides some insight as to what makes one problem harder than another.

Keywords: Analogy; comparison; qualitative representations; spatial reasoning; sketch perception.

Introduction

Understanding how high-level visual properties, such as geometric relationships, are computed is a central problem in perception. One method of exploring what properties are computed is the oddity task. That is, participants are given an array of stimuli, and told to pick the one that is "different" or "odd". If people can do it easily, then they must be computing the property that distinguishes one stimulus from the others, assuming no confounds of course. Dehaene et al. (2006) used the visual oddity task to investigate perception of simple geometric shapes across different cultures. Participants were shown a series of arrays containing six similar images (Figure 1). They were asked to pick out the image that did not fit with the other five. The participant pool included both Americans and Mundurukú, a South American indigenous group, and both children and adults. One finding was that certain problems were much harder than others, for all participant groups. By looking at what makes some problems harder than others, we can gain insight into both what visual properties people tend to compute, and also how they detect oddities. For this paper, we focus entirely on their results for American children, aged 8 to 13. Figure 1 shows their accuracy on a subset of the problems. There are 45 problems in all.

This paper describes a computational model of the visual oddity task. The two key ideas are: (1) Qualitative spatial relations play an important role in much of visual processing (Forbus, Ferguson, & Usher, 2001). Thus, when participants are given a visual array such as the ones used in this study, we propose that they construct a qualitative representation of each image in the array. We model this in our simulation by automatically generating representations with our sketch understanding system, CogSketch[1] (Forbus et al., 2008). (2) Qualitative spatial representations are compared via structure-mapping (Gentner, 1983). In structure-mapping, relational representations are compared by aligning their common structure, which highlights common features and makes it easier to spot the image that lacks those features (cf. Markman & Gentner, 1996). The visual oddity task is difficult because common features must be identified across multiple stimuli. We use analogical generalization to achieve a similar highlighting effect, as explained below.

The combination of automatically generated qualitative visual representations and structure-mapping has been used to model several spatial tasks, including answering geometric Miller Analogy Test questions (Tomai et al., 2005), solving a subset of the Raven's Progressive Matrices, a visual intelligence test (Lovett, Forbus, & Usher, 2007); and making same-different judgments (Lovett, Gentner, & Forbus, 2006). However, none of these tasks offer as much discriminatory power in terms of testing for the presence or absence of particular visual properties.

We begin by briefly reviewing the Structure-Mapping Engine (SME), since it plays a key role in multiple stages of the model. Next we outline our qualitative spatial representations, including how we represent properties of both edges and shapes. Then we describe how comparisons and analogical generalization are used to perform the task. Initial simulation results are discussed, including some predictions from the model. We close with future work.

The Structure-Mapping Engine

SME (Falkenhainer et al. 1986) is a computational model of comparison. Structured, relational descriptions are assumed, including higher-order relations that connect and constrain lower-order relations. Given two descriptions, a base and a target, SME computes one or more *mappings*. A mapping consists of (1) a set of *correspondences*, which indicate what goes with what between the two descriptions, (2) a set of *candidate inferences* that represent conjectures about the target, using the correspondences and unmapped structure in the base, and (3) a *structural evaluation score*, a numerical estimate of overall similarity. SME prefers mappings with high *systematicity*, where connected relational structure, especially with higher-order relations, is mapped.

[1] http://spatialintelligence.org/projects/cogsketch_index.html

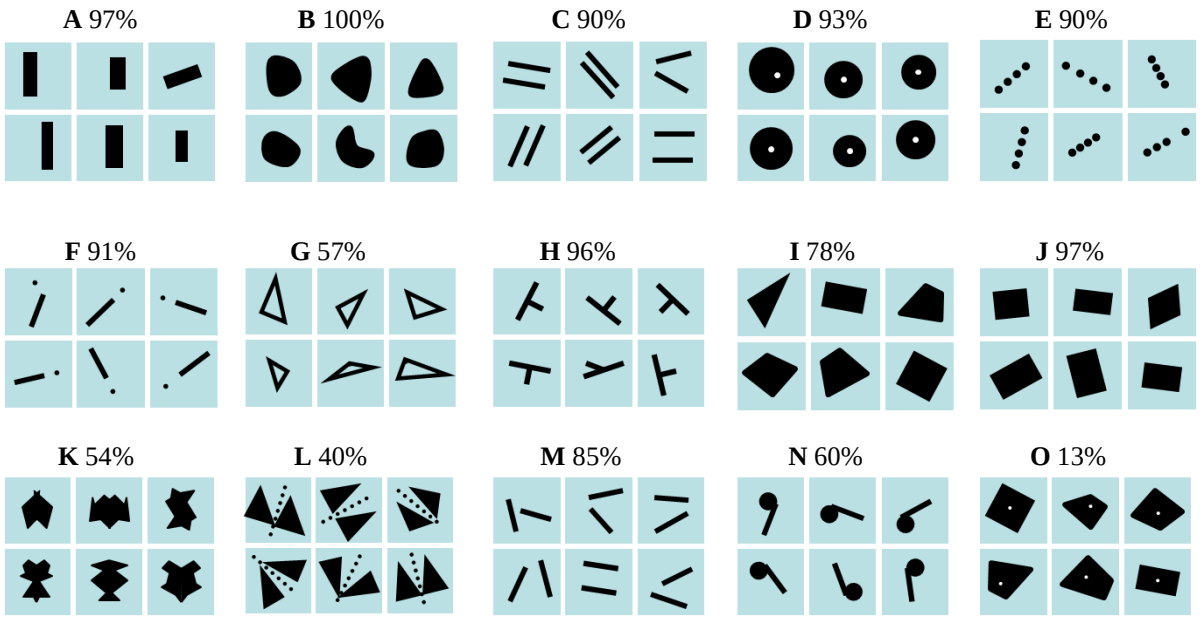

Figure 1. A subset of the 45 problems used by Dehaene et al. (2006). Accuracy is for Americans, aged 8-13.

Qualitative Representation

We believe qualitative relationships are important for comparison tasks because they are much less susceptible to noise than quantitative representations. For example, in comparing two drawings of a face, the important features are qualitative: each face contains an outer ellipse (the head) containing two horizontally aligned circles (the eyes) above two other ellipses (the nose and mouth). Most quantitative data, such as the size of each shape and the orientation of the edges, are not stable across small changes in a drawing. Ideally, qualitative representations should encode what Biederman (1987) calls *nonaccidental properties*. Parallel edges are an example of a nonaccidental property because the range of possible orientations means that edges are unlikely to be parallel by chance. Similarly, two edges are unlikely to be connected by chance.

There is psychological evidence that a number of the features tested for by Dehaene et al. correspond to qualitative attributes and relations encoded by humans. The well established "oblique effect" (Apelle, 1972) shows that humans have a preference for objects aligned with the vertical or horizontal axis (see Figure 1, Problem A). Adults, and even infants as young as five months, can easily distinguish convex and concave objects (Bhatt et al., 2006) (see Problem B), and the salience of parallel lines has been shown in children as young as three (Abravanel, 1977) (see Problem C). Huttenlocher et al. (1991) demonstrated that individuals appear to divide a circle into four quadrants and qualitatively encode which quadrant a dot lies in; it might follow that individuals also encode a relation for cases where the dot lies directly in the circle's center, where the four quadrants meet (Problem D).

Other problems might be solved via qualitative relations based on Gestalt grouping rules (Wertheimer, 1924/1950). For example, grouping by proximity would result in qualitative differences between a single group of proximal dots and two groups of dots, as in Problem E, and the good continuation rule might cause individuals to encode a qualitative relation for a dot that lies along the continuation of a line in Problem F.

Modeling Representation

It has been argued (e.g., Palmer, 1977) that people construct hierarchical spatial representations. Our model constructs qualitative spatial representations at two levels: the edge level and the shape level. The edge level consists of edges, attributes of edges, and relations between edges. The shape level is similar, but for entire shapes. Comparisons are done with either the edge level or the shape level, never both.

Our model generates representations based on *glyphs*, objects that have been sketched in CogSketch. The model assumes the user has sketched each object as a separate glyph. Thus, it does not need to segment a sketch into objects. Each object, or shape, is automatically segmented into edges, using maximal derivatives of the curvature to identify corners between edges along the outline of a glyph. For example, a square would be segmented into four edges, while a circle consists of only a single, elliptical edge.

Each shape has its own edge representation. Table 1 summarizes qualitative edge attributes and relations. Many relations are based on corners between edges. The other relations can only hold for edges that are not connected by a corner along the shape.

Table 2 summarizes attributes and relations for shapes. *Empty/filled* is a simplification of shape color; it refers to whether the shape has any fill color. *Frame-of-Reference*

Edge Attributes	*Edge Relations*
• Straight/Curved/Ellipse • Axis-aligned (horizontal or vertical) • Short/Med/Long (relative length)	• Concave/convex corner • Perpendicular corner • Edges-same-length corner • Intersecting • Parallel • Perpendicular

Table 1. Qualitative vocabulary for edges

relations describe where a smaller shape is located inside a larger, symmetric shape (i.e., a circle). The location of the inner shape is described in terms of quadrants, and whether or not the inner shape is at the central point where the axes of symmetry meet. Currently, grouping by proximity is only implemented for circles.

Line/Line and Line/Point relations apply only to special shape types. Line/Line relations are for shapes that are simple, straight lines (thus these relations are a subset of the edge relations). Line/Point relations are for when a small circle lies near a line. The *centered-on* relation applies when the circle lies at the center of the line. This relation is essentially a special case of the frame-of-reference relation for a dot lying at the center of a circle.

A few shape features require an extra step to compute: *axes of symmetry*, *same-shape*, *rotation-between*, and *reflection-between*. These features can only be computed by using SME to compare shapes' edge representations (Lovett et al. 2007). Axes of symmetry are computed using MAGI (Ferguson, 1994), an extension of SME that compares a representation to itself to look for symmetry. Same-shape is identified by using SME to compare two shapes' edges, using the correspondences to find corresponding edges, and then comparing the edges quantitatively to detect whether the edge mapping represents a rotation or reflection between two instances of the same shape.

Analogical Generalization

Most of the 45 problems can be solved by identifying a qualitative feature that five of the images possess and one image lacks. In a few cases, a problem appears to require noticing that one image possesses a feature that the other five lack, such as parallel lines (Figure 1, Problem M). In either case, multiple images must be compared to identify common features. In essence, participants must build a generalization from the objects. We perform generalization using SEQL (Kuehne et al., 2000), a model of analogical generalization built upon SME. SEQL is based upon the idea that individuals learn generalizations for categories through a process of *progressive abstraction* (Gentner & Loewenstein, 2002), in which instances of a category are compared and the commonalities are abstracted out as a direct result of the comparison.

SEQL uses SME to compare structural representations of objects. When it finds two objects that are sufficiently similar, it constructs a generalization of the objects. A generalization consists of only those elements that correspond with each other in SME's mapping between the objects. Thus, elements found in only one of the two objects are abstracted out of the generalization. The generalization can then be compared to new objects. Each time an object is added to the generalization, the generalization is refined to contain only those elements that align with every object that is part of that generalization.

Oddity Task Model

Our model is based on the following claims about human performance on the oddity task:

1) Humans compute qualitative, relational representations of visual scenes, which they use to solve spatial tasks.
2) Spatial representations for a given operation will always be at either the edge level or the shape level; these two representational levels will not be combined.
3) Representations will be compared via structure-mapping (SME).
4) Analogical generalization (SEQL) will be used to build up a representation of what is common across an array of images in the oddity task.
5) Individual images can be compared to the generalization, and the odd image out should be the one that is noticeably less similar.

In this section, we will describe a task model which is based on these five claims. In order for us to build an operational model, we had to make a number of assumptions beyond these key claims. Some of these assumptions may not be true of human performance, or may not generalize to all other stimuli. However, we believe the overall framework of the model is sound, and we believe the results support the model.

Modeling the Process

Our model attempts to pick out the image that does not belong by performing a series of trial runs. In each trial, the system constructs a generalization from half of the images in the array (either the top half or the bottom half). This generalization represents what is common across all three images. For example, consider the right-angled triangle

Shape Attributes	*Shape Relations*
• Closed shape • Convex shape • Circle shape • Empty/Filled • Axis (Symmetric, Vertical, and/or Horizontal)	• Right-of/Above (relative position) • Containment • Frame-of-Reference • Shape-proximity-group • Same-shape • Rotation-between • Reflection-between
Line-Line Relations	*Line-Point Relations*
• Intersecting • Parallel • Perpendicular	• Intersecting • Colinear • Centered-On

Table 2. Qualitative vocabulary for shapes

problem (Figure 1, Problem G). The generalization built from the three top images will describe three connected edges, with two of the edges being perpendicular. In the leftmost top image, the two perpendicular edges are of different lengths, but this relation will have been abstracted out because it is not common to all three images.

The generalization is then compared to each of the other three images, using SME. The model examines the similarity scores for the three images, looking for a particular pattern of results: two of the images should be quite similar to the generalization, while the third image, lacking a key feature, should be less similar. In this case, the lower middle triangle will be less similar to the generalization because it lacks a right angle.

Similarity is based on SME's structural evaluation score, but it must be normalized. There are two different ways to normalize it: Similarity scores can be normalized based only on the size of the generalization (*gen-normalized*), which measures how much of the generalization is present in the image being compared. This measure is ideal for noticing whether an image lacks some feature of the generalization.

Alternatively, similarity scores can be normalized based on both the size of the generalization and the size of the image's representation (*fully-normalized*). This score measures both how much of the generalization is present in the image and how much of the image is present in the generalization. While more complex than gen-normalized scores, fully-normalized scores are necessary for noticing an oddity that possesses an extra qualitative feature that the other images lack. For example, it allows the model to pick out the image with parallel lines from the other five images without parallel lines.

Controlling the Processing

In each trial run, the model must make three choices. The first is whether to generalize from the top three images or the bottom three images. The second is whether to use gen-normalized or fully-normalized similarity scores. The third is whether to use edge representations or shape representations. These choices are made via the following simple control mechanism: (1) To ensure that the results are not dependent on the order of the images in the array, trial runs are attempted in pairs, one based on generalizing from the top three images and one based on generalizing from the bottom three images. (2) Because the gen-normalized similarity score is simpler, it is always attempted first. (3) The model chooses whether to use edge or shape representations based on the makeup of the first image. If the image contains multiple shapes, or if the image contains an elliptical shape consisting of only a single edge (e.g., a circle), then a shape representation is used. Otherwise, an edge representation is used. Note, however, that an edge representation will be quickly abandoned if it is impossible to find a good generalization across images, as indicated by different images having different numbers of edges.

After the initial pair of trials is run, the model looks for a *sufficient* candidate. Recall that each trial run produces three similarity scores for the three images compared. A sufficient candidate is chosen when the lowest-scoring image has a similarity score noticeably lower than the other two (< 95% of the second lowest-scoring image) and the other two images are reasonably similar to the generalization (normalized score > 55%).

When a sufficient candidate is not found, the model attempts additional trial runs. (1) If the model was previously run using edge representations, it will try using shape representations. (2) The model will try using a fully-normalized similarity score, to see if the oddity possesses an extra feature. At this point, if no sufficient candidate has been identified, the model gives up. We do not allow the model to guess randomly, as people sometimes do.

Predictions

This model suggests five factors that ought to contribute to the difficulty of a problem:

1. Feature computability. The first requirement for identifying a common feature is being able to compute it. Individuals who are unable to compute the key feature cannot solve the problem. Problem O, for example, requires participants to determine whether the dot falls at the intersection of the quadrilateral's axes. An inability to compute this feature would contribute to this being one of the hardest problems.

2. Feature salience. Salience here means the likelihood that participants will encode a particular feature. There are far more possible visual properties that could be computed than finite attention and resources permit to actually be computed. A low-salience feature might not be computed at first, and only generated in a later trial run when the most salient properties don't lead to an answer. Our model predicts that when images have multiple shapes, shape features will be much more salient than edge features, whereas when there is only a single shape, edge features will be more salient. This could explain the difficulty of problems such as K, which rests on the symmetry of the shape, rather than any features of individual edges.

3. Feature representation strength. Because of SME's systematicity preference, it assigns higher similarity scores to correspondences that support large relational structures. Therefore, absence of features represented by higher-order relations should be easier to spot, since they will influence similarity scores more. Similarly, if a feature is represented as multiple relations, its absence will be easier to spot than if it were represented by only a single relation. Of course, representation strength is relative; in a sparse representation, the absence of even a single attribute may be easy to spot. This could explain why, for example, participants are much better at solving a problem based on two perpendicular lines than they are at solving a problem based on a right corner in a triangle (Problems H and G). The representation of two perpendicular lines would be much sparser than the representation of a right triangle, so the relative strength of the relation specifying that two edges are perpendicular would increase.

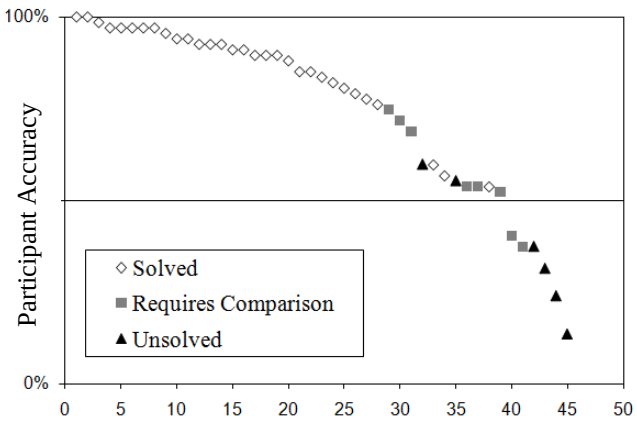

Figure 2: Performance by our model on the 45 problems (ranked by difficulty for human participants)

4. Feature presence versus feature absence. Because the model uses the gen-normalized similarity score before the fully-normalized similarity score, it solves problems in which the oddity lacks a feature more quickly than when the oddity possesses an added feature. Thus, the model predicts that participants should be faster and more accurate when solving problems where the oddity lacks the feature. Unfortunately, it is difficult to evaluate this prediction based on the current data, as there are only a few problems in which the oddity has an added feature. The one case where an oddity has an added feature is in one problem and lacks that same feature in another involves parallel lines, and participants performed similarly on both problems (Problems C and M). However, this may have been because both problems were quite easy.

5. Alignability of images. Participants should find a problem more difficult if it is harder to align the five common images. This is because (a) there will be less structural support for the initial generalizations and (b) the similarity scores between any of the images and the generalization will be lower. For example, participants had more difficult picking a triangle out of quadrilaterals (Problem I) than picking a parallelogram out of rectangles (Problem J). Even though a triangle is easier to distinguish from quadrilaterals, all the quadrilaterals were different from each other, thus making it harder to align them with each other to determine what common feature they possessed that the triangle lacked.

Evaluation

We evaluated our model by running it on all 45 problems from the original study (Dehaene et al., 2006). The original stimuli, which had been drawn in PowerPoint, were copied and pasted into CogSketch. Of the 45 problems, four were touched up in PowerPoint to ease the transition—lines or polygons that had been drawn as separate parts and then grouped together were redrawn as a single shape. In addition, five problems were modified after being pasted into CogSketch. In all five cases, we removed simple edges which had been added to the images of the problem to help illustrate an angle or reflection participants were meant to attend to (e.g., Problem L). Because the model was not able to understand the message these lines were meant to convey, they would have served only as distracters. Aside from the changes to these nine problems, no changes were made to the stimuli which had been run on human participants.

CogSketch treats each PowerPoint object (line, polyline, or polygon) as a separate glyph and thus a separate object. After the problems were pasted into CogSketch, it computed the spatial relations between each edge in an object, producing the edge representations for a problem. It also computed object attributes and relations between objects in each image of a problem, producing the shape representations for a problem. The model then attempted to solve the problem using the method described above.

Results

Given the 45 problems, our model successfully solved 39 problems. Note that chance performance on the task would be solving 7.5 problems.

We ranked the problems based on the difficulty that the children had solving them, with 45 being the hardest. Of the six problems missed by our model, four were also the four hardest problems for the children. The other two were among the harder problems, at positions 32/45 and 35/45. Thus the average difficulty rank of the problems missed was 40.2/45. Figure 2 shows the difficulty of the problems the model was unable to solve. The hardest problem for children was Problem O (in Figure 1), in which the key feature was whether a dot lied along the axes between the corners of a quadrilateral. Our model simply does not compute this feature, nor do the children, we believe, as they scored below chance on this problem.

The other five problems missed by our model all required that participants either encode a quantitative feature for each image or directly compare shapes between images. For example, consider Problem N, in which the key feature was the position of the circle relative to the line. It appears that this problem could only be solved by comparing the shapes in pairs of images and mentally rotating them to determine whether they align. Our model compares shapes and looks for rotations within a single image, but not across different images of the array.

These results suggest that problems requiring comparing shapes across two separate images were particularly difficult, given that both the model and participants had trouble solving these problems. This led us to ask whether problems which required comparing shapes within a single image would also be difficult. We ran a second evaluation in which our model did not compute any of the shape comparisons—these included rotations and reflections between shapes, as well as axes of symmetry within a shape that could only be computed by comparing the shape to itself with MAGI. See Figure 2 again for the difficulty of the problems the model was unable to solve without shape comparisons. These eight problems, along with the six the model failed to solve initially, make up 14 of the 17 hardest

problems for children. Thus, they nearly perfectly match the hardest third of the problem set.

Conclusions and Future Work

We believe the results described above provide strong support for our model of the visual oddity task. Qualitative spatial representations can be used with structure mapping and analogical generalization to solve nearly all of the problems from the original Dehaene et al. (2006) study. The problems on which the model fails are among the hardest problems for human participants. Furthermore, while edge representations are sometimes used to identify relations between shapes (such as rotations and reflections), the overall comparison mechanism is always run on either edge representations or shape representations. Thus, the model suggests that individuals do not need to represent edges and shapes simultaneously while making comparisons.

Feature computability and salience seem to be the two factors contributing the most to problem difficulty. The model failed on the problems for which it was unable to compute the key feature, such as the relative position of a line and a circle once the shapes have been rotated to the same orientation. Moreover, the model correctly predicted that people would have difficulty with other problems in which the key feature could only be computed by comparing the shapes within one image of an array. In other words, the current results suggest that people often fail to compare individual shapes before comparing the images themselves to solve these problems.

One line of investigation for the future concerns sharpening the model's explanation of problem difficulty, by conducting a more detailed analysis of the model's output and its relationship to human results. Several extensions of the model are also intriguing, e.g., modeling feature salience via a probabilistic representation scheme.

Our long-term goal is to develop a general model of human qualitative spatial representation. Each spatial task which we have modeled (e.g., Tomai et al., 2005; Lovett et al., 2006; Lovett et al., 2007) puts constraints on the representation that may be used to solve that particular task. A spatial representation scheme that works across all of these tasks will have much stronger support as a model of human spatial representation.

Acknowledgements

This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC).

References

Abravanel, E. (1977). The figural simplicity of parallel lines. *Child Development, 48*(2), 708-710.

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animal. *Psychological Bulletin, 78*, 266-278.

Bhatt, R., Hayden, A., Reed, A., Bertin, E., & Joseph, J. (2006). Infants' perception of information along object boundaries: Concavities versus convexities. *Experimental Child Psychology, 94*, 91-113.

Biederman, I. (1987). Recognition-by-Components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science, 311*, 381-384.

Falkenhainer, B., Forbus, K., & Gentner, D. (1986). The Structure-Mapping engine. *Proceedings of the Fifth National Conference on Artificial Intelligence*.

Ferguson, R. W. (1994). MAGI: Analogy-based encoding using regularity and symmetry. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*.

Forbus, K., Usher, J., Lovett, A., & Wetzel, J. (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. *Proceedings of the Eurographics Workshop on Sketch-Based Interfaces and Modeling*.

Forbus, K., Ferguson R., & Usher, J. (2001). Towards a computational model of sketching. *Proceedings of the 2001 Conference on Intelligent User Interfaces.*

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science 7*(2), 155-170.

Gentner, D., and Loewenstein, J. (2002). Relational language and relational thought. In E. Amsel and J. P. Byrnes (Eds.), *Language, Literacy, and Cognitive Development: The Development and Consequences of Symbolic Communication.* Lawrence Erlbaum Associates.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating location. *Psychological Review, 98*(3), 352-376.

Kuehne, S., Forbus, K., Gentner, D., and Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*.

Lovett, A., Forbus, K. & Usher, J. (2007). Analogy with qualitative spatial representations can simulate solving Raven's Progressive Matrices. *Proceedings of the 29th Annual Conference of the Cognitive Society*.

Lovett, A., Gentner, D., and Forbus, K. (2006). Simulating time-course phenomena in perceptual similarity via incremental encoding. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*(2), 235-249.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9*(4), 441-474.

Tomai, E., Lovett, A., Forbus, K., & Usher, J. (2005). A structure mapping model for solving geometric analogy problems. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Stresa, Italy.

Wertheimer, M. (1924/1950). Gestalt theory. In W. D. Ellis (Ed.), *A Sourcebook of Gestalt Psychology* (pp. 1-11). New York: The Humanities Press.