Shape is like *Space*: Modeling Shape Representation as a Set of Qualitative Spatial Relations

Andrew Lovett **Kenneth Forbus**

Qualitative Reasoning Group, Northwestern University
andrew-lovett@northwestern.edu
forbus@northwestern.edu

Abstract

Representing and comparing two-dimensional shapes is an important problem. Our hypothesis about human representations is that that people utilize two representations of shape: an abstract, *qualitative* representation of the spatial relations between the shape's parts, and a detailed, *quantitative* representation. The advantage of relational, qualitative representations is that they facilitate shape comparison: two shapes can be compared via structural alignment processes which have been used to model similarity and analogy more broadly. This comparison process plays an important role in determining when two objects share the same shape, or in identifying transformations (rotations and reflections) between two shapes. Based on our hypothesis, we have built a computational model which automatically constructs both qualitative and quantitative representations and uses them to compare two-dimensional shapes in visual scenes. We demonstrate the effectiveness of our model by summarizing a series of studies which have simulated human spatial reasoning.

Introduction

Humans possess an impressive capacity for identifying, representing, and reasoning over the spatial relationships between objects. For example, consider a geometric analogy problem (Figure 1A), in which an individual must answer the question "A is to B as C is to…?" Answering this problem requires recognizing three things: 1) in the mapping from A to B there is a reversal of vertical positions, 2) in the mapping from C to 3 there is a reversal of horizontal positions, and 3) both of these mappings are instances of some more abstract "switching places" relation. However, people can answer problems like this one quickly and accurately (Lovett et al., 2009b).

In addition to reasoning intelligently about the relations between objects' *locations*, we are also quite skilled at reasoning about the relations between objects' *shapes*. For example, Figure 1B shows a problem that can only be solved if one recognizes that the triangle shape is being flipped or rotated in the mapping between A and B. Our facility in solving this problem (which people actually answer about twice as fast as the previous problem) demonstrates our ability to integrate both spatial and shape relations in solving problems like geometric analogies.

Computing relations between shapes—determining that they are identical, or that one is a rotation or reflection of the other—is a difficult problem. In the past, there have been heated arguments over the nature of people's shape representations, particularly whether people use *orientation-invariant* representations (Biederman 1987; Biederman and Gerhardstein 1993) that automatically align during comparison or *orientation-specific* representations (e.g., Tarr et al. 1997) that must be mentally rotated during

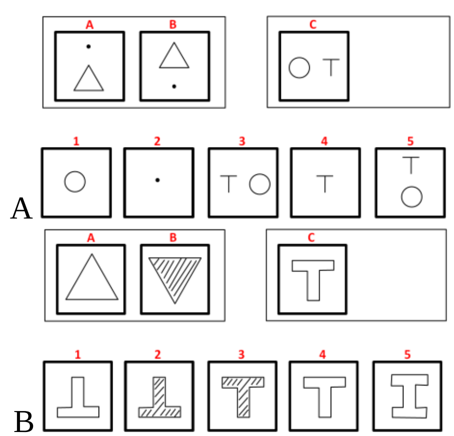


Figure 1. Two geometric analogy problems from (Evans, 1968; Lovett et al. 2009b)

comparison. However, we believe the most parsimonious approach to modeling shape cognition is to assume that there are no special representations or processes for dealing with shapes. Rather, we believe, people represent and reason about shapes in the same way that they represent and reason about space, only on a smaller scale.

In particular, we make the following key claims about shape representation and comparison:

1) People rely on two types of shape representations: an abstract, qualitative, orientation-invariant representation and a detailed, quantitative, orientation-specific representation (e.g., Hummel 2001).

2) Qualitative representations of shape, like qualitative representations of space, are structured representations (Biederman 1987) describing the relations between a set of elements. For shape representations, these elements are the parts of the shape (e.g., the edges of a line drawing).

3) Qualitative shape representations can be compared via structure-mapping (Gentner 1983), a process of aligning their common relational structure. Structure-mapping is believed to play a role in a variety of domains, from abstract analogies to visual scene comparison (Lovett et al. 2009a).

4) Performing a more exact comparison or determining a specific transformation between two shapes requires comparing their quantitative representations. This is a two-step process in which people first use structure-mapping on the *qualitative* representation to identify the corresponding parts in the shapes and then mentally transform the *quantitative* representations to align the corresponding parts.

In this paper, we describe a computational model of shape representation and comparison based on the claims given above. We begin by summarizing work in psychology that supports the claims. We then describe our models of shape representation and comparison. Finally, we review several spatial reasoning tasks in which we have used our model to achieve human-level performance.

Background

Several areas of psychological research have informed our model. We consider each in turn.

Perceptual Organization

Perceptual organization is concerned with how people divide a visual scene into a coherent set of objects. It is generally assumed (e.g., Ullman 1987; Palmer and Rock 1994) that bottom-up processes create a set of *entry-level units*. These initial elements can then be grouped into larger objects or parsed into smaller objects.

A conclusion one may draw from these ideas is that the distinction between the objects in a scene and the parts in an individual object is a difference of degree only; one of these may be represented and reasoned over as easily as the other. Indeed, Palmer (1977) suggested that people possess hierarchical representations of visual space. At any level of representation, there are a set of *structural units* (SUs), each of which possesses a set of relations to other SUs at its level. Zooming in on any single SU, one can identify the set of SUs and relations that make it up.

This suggests that, just as space can be represented as a set of objects and relations between objects, the shape of an individual object can be represented as a set of parts and relations between parts (e.g., Biederman 1987). Similarly, just as visual scenes can be compared by aligning their common structure and identifying the corresponding objects within them (Lovett et al. 2009a), shapes can be compared by aligning their common structure and identifying the correspond parts within them.

Spatial Relations

If shape, like space, is represented as a set of relations, what should the nature of those relations be? We believe there is strong evidence that people typically encode *qualitative*, or categorical, relations between object parts.

For example, let us consider the lines that make up two-dimensional shapes. There is evidence that people encode qualitative relations for both the relative orientation of two lines and the type of connection between two lines. For relative orientation, people show a strong sensitivity to both parallel lines (Abravanel 1973), and perpendicular lines (Chen and Levi 1996); however, the sensitivity to perpendicular lines may appear only when one line is vertical and one line is horizontal, perhaps due to alignment of a line with the x- and y-axes being a salient qualitative feature.

For connections between edges, there is a great deal of evidence that people are sensitive to concave corners between the edges of a shape (e.g., Ferguson et al. 1996; Elder and Zucker 1993). We believe it is likely that people make a qualitative distinction between concave and convex corners within a shape.

Object Recognition

There has been a lengthy, ongoing debate about how objects' shapes are represented and compared for recognition. One major question is whether representations are orientation-invariant or orientation-specific. If representations are orientation-invariant, then the process of recognizing involves simply encoding an object's representation and directly comparing it to other representations until a match is found. On the other hand, if representations are orientation-specific, then the process of recognizing depends critically on the orientation at which an object is perceived. If that orientation is an unfamiliar one, the individual may need to mentally transform the object's representation to line it up with a more familiar orientation. Thus, recognizing objects at unfamiliar orientations should be at least slower, and possibly more error-prone.

Despite the clear, differing predictions these two viewpoints make, determining which is correct has proven

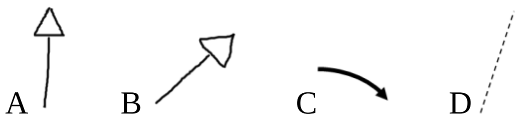


Figure 2. A,B: Two arrow shapes. C: A rotation between the shapes. D: An axis of reflection between the shapes.

difficult. In some studies (e.g., Biederman and Gerhardstein 1993), individuals have been able to quickly recognize newly learned objects regardless of orientation. In others (e.g., Tarr et al. 1997), unfamiliar orientations have hampered the recognition process. Thus, it appears that orientation-invariant recognition works sometimes, depending on the specifics of the task.

We believe the best explanation of such conflicting results is that people possess two types of shape representations. The first is an abstract, orientation-invariant representation describing the qualitative spatial relations between parts of the shape, as suggested above. The second is a concrete, orientation-specific representation describing the shape's complete details. The exact form of this second representation is less clear. Whereas the orientation-*invariant* representations can be quickly compared to compute a rough measure of similarity between two shapes, the orientation-*specific* representations must be transformed and compared to determine for certain whether two shapes are identical.

Mental Rotation

Mental rotation can be seen as a particularly difficult special case of object recognition. In mental rotation tasks (Shepard and Metzler 1971; Shepard and Cooper 1982), individuals are shown two shapes (either two- or three-dimensional) and asked to determine whether a rotation of one would produce the other. To make this more difficult, the false trials are typically created by reflecting and then rotating one shape to create the other. Thus, even on the false trials the shapes are quite similar.

The typical finding in mental rotation tasks is that the time to determine that one shape is a rotation of the other is proportionate to the degrees of rotation between them. This suggests that individuals are applying some kind of analog rotation to their representation of one of the shapes to line it up with the other.

At first glance, the mental rotation studies simply seem like more evidence for an orientation-specific representation of shape. However, there is a deeper question of what guides the mental rotation process. In most cases, the reaction time on this task is proportionate to the degrees of rotation along the *shortest* possible rotation between the two shapes (Shepard and Cooper 1982). This leads to the question of how people know which way to rotate one shape to line it up with the other.

We believe the mental rotation results can be best explained via a two-step process. This begins with a quick comparison of the qualitative, structural, rotation-invariant representations. By aligning the common structure in the representations being compared (Gentner 1983), one can quickly identify the corresponding parts in the two shapes. An understanding of these correspondences can then guide the comparison of the orientation-specific representations.

For example, consider the two arrow shapes in Figure 2. An individual might compare them in the following manner: 1) Compare the structured, qualitative representations to determine that the stems of the two arrows correspond to each other. 2) Compare the stems to quickly determine that there is a 45-degree rotation between them (Figure 2C). 3) Mentally apply the 45-degree rotation to the detailed, quantitative representations, and see if it results in those representations lining up with each other. This third step would be the only one whose speed would depend on the degrees of rotation between the two shapes.

A Model of Shape Representation

Our model of shape representation is built into the CogSketch sketch understanding system (Forbus et al. 2008). CogSketch takes a set of objects, or *glyphs*, drawn by the user and automatically computes qualitative spatial relations between them. Our model does the same thing for the edges making up a single glyph in CogSketch. Shape representations are computed via a three-step process:

1) Identify the edges and junctions in the glyph.

2) Group edges into cycles representing closed shapes.

3) Generate a qualitative description of the spatial relations between the edges.

This qualitative description is our model's orientation-*invariant* representation. We have no strong theoretical commitment on the form of the orientation-*specific* representation. In our model, it is simply the set of edges and their orientations, along with the glyph's overall size.

As we describe the three steps of this process in detail, we use Figure 2A, the arrow shape, as a running example.

Identifying Edges and Junctions

Our model takes as input a set of *polylines*, lists of points describing lines drawn by a CogSketch user. Rather than assuming the user drew one line for each edge of the shape—e.g., four lines for the arrow shape—the model first merges those polylines whose endpoints touch and then begins segmenting the polylines into edges.

Several researchers (e.g., Clowes 1971; Bierderman 1987) have suggested that the human visual system makes use of *junctions*, locations where two or more edges meet, when building up a shape representation. Our model relies on junctions to parse a shape into edges, identifying junctions in the shape and then treating edges as maximally long lists of adjacent points between junctions.

The model utilizes two approaches for identifying junctions between edges. Firstly, any place three or more lines meet is a junction (Figure 3). Typically, i.e., in fork- and arrow-junctions, it is assumed that these lines are

L-junct Fork-junct Arrow-junct T-junct X-junct

Figure 3. Common junction types (most from Clowes, 1971)

distinct edges. However, there are two exceptions: T-junctions, where one edge bisects another edge, and X-junctions, where two edges intersect each other. In Figure 2A, there is a T-junction where the stem of the arrow bisects one of the edges of the head.

The other approach is used to identify L-junctions, where two edges meet. L-junctions are identified based on a local discontinuity in the curvature, accompanied by a larger-scale change in orientation (see Lovett et al., 2009b for details). In Figure 2A, there are three L-junctions between the edges of the arrow's head.

Grouping Edges into Cycles

Closure, i.e., a cycle of edges that together form a closed shape, is one of the most important Gestalt grouping rules in perceptual organization (Rock & Palmer, 1990). In addition, it is needed to identify concavities (Elder & Zucker, 1993), an important and salient relationship between edges in a shape.

Our model identifies edge cycles via the simple expedient of searching exhaustively through the network of edges connected by junctions. While we doubt this is the approach people use, it is sufficient for our purposes. Once identified, an edge cycle is represented by a list of the edges making up the cycle, in clockwise order. This makes it easy to enumerate the corners between edges in the cycle and to classify each corner as *convex* or *concave*, depending on whether it points into or out of the cycle. In Figure 2A, the model would identify a single edge cycle with three edges and three convex corners.

Generating Qualitative Descriptions

Given a list of edges, junctions between edges, and edge cycles, the model generates a qualitative, orientation-invariant representation of the relations between the edges. This representation is based upon the tenets of structure-mapping (Gentner, 1983) a model of analogical and relational comparison which claims that we compare relational descriptions by aligning their common structure. Relational descriptions consist of entities, attributes of entities, and relations between entities. There can also be higher-order relations between other, lower-order relations. According to structure-mapping, people generally try to find the most *systematic* matches between two cases, i.e., the matches with the deepest common structure. Thus, higher-order relations play a key role in the mapping process and are therefore a particularly important part of the shape representation.

In our representations, the entities refer to edges of the shape. The full set of attributes and relations is given in Figure 4. We now describe these in turn.

Attributes describe features of individual edges. Every edge in a shape is a **PerceptualEdge.** Each edge is further assigned an attribute specifying whether it is **straight**, **curved**, or **elliptical**. An **elliptical** edge is an edge that closes on itself, such as a circle. Each edge is also assigned a length attribute based on its length relative to the shape's longest edge. Finally, straight edges can be classified as **axisAligned**, meaning they are aligned with the x- or y-axes. This attribute is the only term in the representations that contains some orientation-specific information. However, as it is a low-level attribute describing a single edge, it plays a small role in the representation.

Simple edge relations are low-level relations describing relationships between pairs of edges. The first two, **edgesPerpendicular** and **edgesParallel**, describe the relative orientations of two edges. **edgesCollinear** describes a special case of parallel edges, in which the edges are also collinear. The other relations describe pairs of edges that meet at different types of junctions. **elementsConnected** describes two edges that meet at an L-junction, **elementsIntersect** describes two edges that intersect at an X-junction, and **elementIntersects** describes one edge that bisects another at a T-junction.

The final set of relations describe relationships between edges in an edge cycle. Firstly, all corners between edges are classified as **convex** or **concave**. **cycleAdjacentAngles** is a higher-order relation describing consecutive pairs of corners along the cycle. **perpendicularCorner** and **equalLengthEdgesCorner** are higher-order attributes describing features of a corner. Finally, **parallelEdgeRelation** and **collinearEdgeRelation** are higher-order attributes for pairs of edges in a cycle. These are included to give parallel and collinear edges the same level of structural depth as perpendicular edges in the shape representation.

Figure 5 shows the representation the model generates for the arrow shape.

Attributes	Simple Edge Relations	Edge Cycle Relations
PerceptualEdge	edgesPerpendicular	convexAngleBetweenEdges / concaveAngleBetweenEdges
StraightEdge/CurvedEdge/EllipseEdge	edgesParallel	cycleAdjacentAngles
length(Tiny/Short/Medium/ Long)	edgesCollinear	perpendicularCorner
axisAligned	elementsConnected	equalLengthEdgesCorner
	elementsIntersect	parallelEdgeRelation
	elementIntersects	collinearEdgeRelation

Figure 4. Qualitative vocabulary for shape representation.

A Model of Shape Comparison

We believe that people can perform two types of shape comparisons. The first is a quick estimate of the similarity of two shapes' orientation-invariant representations. The second is a more careful comparison of the orientation-specific representations which may require applying mental transformations. Because the second type of comparison requires first identifying the corresponding edges in the two shapes, both types of comparison require an efficient method for aligning the structured, orientation-invariant representations. Therefore, we begin this section by describing the Structure Mapping Engine (Falkenhainer et al., 1989), our model of structural alignment.

Qualitative Comparison via Structural Alignment

Qualitative, structural representations are compared using the Structure Mapping Engine (SME) (Falkenhainer et al., 1989), a computational model based on Gentner's (1983) structure-mapping theory of analogy. While SME was originally built to model abstract analogical comparison, it has since been used to model human relational comparisons in a number of domains, including concrete spatial representations (Lovett et al., 2009a). Given two structured representations, a *base* and a *target*, SME computes one or more *global mappings* between them. Each mapping consists of: 1) a set of *correspondences* between elements in the base and target; 2) a *structural evaluation score*, an estimate of similarity based on the degree of overlapping structure; and 3) a set of *candidate inferences*, inferences about the target computed based on elements in the base that failed to map to the target.

SME is useful in modeling shape comparisons for two reasons. Firstly, the structural evaluation score can be used as an estimate of the similarity of the shapes being compared. Secondly, the correspondences indicate which edges in the two shapes correspond to each other. These correspondences can be used to guide a comparison of quantitative representations, as described next.

Quantitative Comparison

Quantitative comparison is necessary to determine for certain whether two objects are the same shape. It also allows one to identify transformations between the objects' shapes. Next we describe our model for identifying rotations and reflections between shapes.

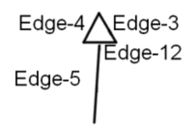
Rotation. Rotations are identified via a three-step process. Firstly, the shapes' qualitative, orientation-invariant representations are compared via SME to identify corresponding edges. The edge correspondences are then carried over to the quantitative, orientation-specific representations. Recall that we model quantitative shape representations as simply a list of edges and their quantitative orientations. We expect that human quantitative representations are considerably richer.

Secondly, the first pair of corresponding edges are compared. Based on their relative orientations, two *candidate rotations* between them are identified. For example, suppose the two arrow shapes in Figure 2 were compared, and the stems of the arrows were identified as corresponding edges. The candidate rotations would be a 45-degree, clockwise rotation and a 135-degree, counter-clockwise rotation.

Thirdly, these candidate rotations are evaluated by considering the relative orientations of the other corresponding edges. For the two arrows, only the 45-degree, clockwise rotation can be consistently applied to all corresponding edges (Figure 2C).

Note that if SME identifies more than one equally valid mapping between the edges of the shapes, this approach can return more than one possible rotation. The model simply sorts these from least to greatest degrees of rotation. In contrast, people generally identify the shortest possible rotation first (Shepard & Cooper, 1982). The model might be made more accurate by adding a small amount of orientation-specific information to the structural representations to bias SME towards returning a mapping representing the shortest possible rotation.

Reflection. Our approach for identifying axial reflections between shapes is similar. There are only two differences: 1) For the qualitative representations, the order of the edges in any edge cycles are reversed for one of the two representations. This is done because axial reflections




```
(PerceptualEdge Edge-3)        (PerceptualEdge Edge-4)
(PerceptualEdge Edge-5)        (PerceptualEdge Edge-12)
(StraightEdge Edge-3)          (StraightEdge Edge-4)
(StraightEdge Edge-5)          (StraightEdge Edge-12)
(lengthShort Edge-3)           (lengthShort Edge-4)
(lengthLong Edge-5)            (lengthShort Edge-12)
(axisAligned Edge-5)           (axisAligned Edge-12)

(elementIntersects Edge-5 Edge-12)

(cycleAdjacentAngles
   (convexAngleBetweenEdges Edge-12 Edge-4)
   (convexAngleBetweenEdges Edge-4 Edge-3))
(cycleAdjacentAngles
   (convexAngleBetweenEdges Edge-3 Edge-12)
   (convexAngleBetweenEdges Edge-12 Edge-4))
(cycleAdjacentAngles
   (convexAngleBetweenEdges Edge-4 Edge-3)
   (convexAngleBetweenEdges Edge-3 Edge-12))

(equalLengthEdgesCorner
   (convexAngleBetweenEdges Edge-12 Edge-4))
(equalLengthEdgesCorner
   (convexAngleBetweenEdges Edge-3 Edge-12))
(equalLengthEdgesCorner
   (convexAngleBetweenEdges Edge-4 Edge-3))
```

Figure 5. Shape representation for the arrow shape.

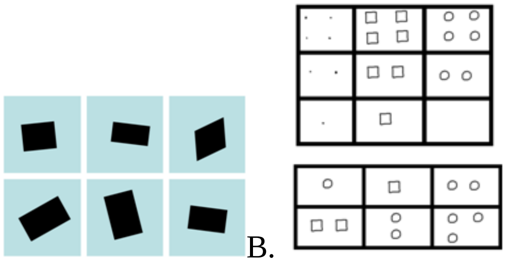


Figure 6. A: Example problem from the Visual Oddity Task: “Pick the image that doesn’t belong.” B: Example problem from Raven’s Progressive Matrices (not an actual test problem): “Pick the image that best completes the matrix.”

result in a reversal of the order of elements in a cycle. 2) In the second and third steps of the process, the model looks for possible axes of reflection between two edges’ orientations, rather than rotations. For the two arrows, there is one possible axis of reflection (Figure 2D).

Studies

We have used our model of shape representation and comparison in a number of studies, although the vocabulary of qualitative spatial relations has evolved over time. These studies include simulations of three spatial problem-solving tasks: geometric analogy (Lovett et al. 2009b; Figure 1), the visual oddity task (Lovett, Lockwood, and Forbus 2008; Figure 6A), and a subset of the Raven’s Progressive Matrices (Lovett, Forbus, and Usher 2007; Figure 6B). All of these tasks require comparing images in order to find patterns in their spatial relations. These tasks also require integrating shape information—shape identity and shape transformations—with spatial information. On all three tasks, our models have achieved adult human-level performance or better, using representations automatically computed from stimuli similar or identical to those presented to human participants.

Conclusion

We have presented a psychologically-motivated computational model of shape representation and comparison. The key insights of this model are that shape, like space, can be represented as a set of qualitative spatial relations between parts, and that shape representations can be compared via structure mapping. While other researchers have explored the first idea before (Biederman, 1987; Hummel, 2001), our model moves beyond their work by automatically building up shape representations that can be used in larger spatial reasoning tasks.

Acknowledgements

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

References

Abravanel, E. 1977. The Figural Simplicity of Parallel Lines. *Child Development* 48(2): 708-710.

Biederman, I. 1987. Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review* 94(2): 115-147.

Biederman, I., and Gerhardstein, P. C. 1993. Recognizing Depth-Rotated Objects: Evidence and Conditions for 3D Viewpoint Invariance. *Journal of Experimental Psychology: Human Perception and Performance* 19: 1162-1182.

Chen, S., and Levi, D. M. 1996. Angle Judgment: Is the Whole the Sum of Its Part? *Vision Research* 36(12): 1721-1735.

Clowes, M. B. 1971. On Seeing Things. *Artificial Intelligence* 2: 79-116.

Elder, J., and Zucker, S. 1993. The Effect of Contour Closure on the Rapid Discrimination of Two-Dimensional Shapes. *Vision Research* 33(7): 981-991.

Evans, T. 1968. A Program for the Solution of Geometric-Analogy Intelligence Test Questions. In M. Minsky (Ed.), *Semantic Information Processing*. Cambridge: MIT Press.

Falkenhainer, B., Forbus, K., and Gentner, D. 1989. The Structure Mapping Engine: Algorithm and Examples. *Artificial Intelligence* 41: 1-63.

Ferguson, R., Aminoff, A., and Gentner, D. 1996. Modeling Qualitative Differences in Symmetry Judgments. In Proceedings of CogSci 96.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., and Wetzel, J. 2008. CogSketch: Open-Domain Sketch Understanding for Cognitive Science Research and for Education. In Proceedings of SBIM ‘08.

Gentner, D. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7: 155-170.

Hummel, J. E. 2001. Complementary Solutions to the Binding Problem in Vision: Implications for Shape Perception and Object Recognition. *Visual Cognition* 8: 489–517.

Lovett, A. Forbus, K., and Usher, J. 2007. Analogy with Qualitative Spatial Representations Can Simulate Solving Raven's Progressive Matrices. In Proceedings of CogSci09.

Lovett, A., Gentner, D., Forbus, K., and Sagi, E. 2009a. Using Analogical Mapping to Simulate Time-Course Phenomena in Perceptual Similarity. *Cognitive Systems Research* 10(3): Special Issue on Analogy: 216-228.

Lovett, A., Lockwood, K., and Forbus, K. 2008. Modeling Cross-Cultural Performance on the Visual Oddity Task. In Proceedings of Spatial Cognition ‘08.

Lovett, A., Tomai, E., Forbus, K., and Usher, J. 2009b. Solving Geometric Analogy Problems through Two-Stage Analogical Mapping. *Cognitive Science* 33(7): 1192-1231.

Palmer, S. E. 1977. Hierarchical Structure in Perceptual Representation. *Cognitive Psychology* 9: 441-474.

Palmer, S., and Rock, I. 1994. Rethinking Perceptual Organization: The Role of Uniform Connectedness. *Psychonomic Bulletin & Review* 1(1): 29-55.

Rock, I., and Palmer, S. 1990. The Legacy of Gestalt Psychology. *Scientific American* 263: 84-90.

Shepard, R. N., and Cooper, L. A. 1982. *Mental Sketches and Their Transformations*. Cambridge: MIT Press.

Shepard, R. N., and Metzler, J. 1971. Mental Rotation of Three-Dimensional Objects. *Science* 171: 701-703.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., and Blanz, V. 1997. To What Extent Do Unique Parts Influence Recognition across Viewpoint? *Psychological Science* 8(4): 282-289.

Ullman, S. 1984. Visual Routines. *Cognition* 18: 97-159.