Modeling Spatial Abstraction during Mental Rotation

Andrew Lovett (andrew@cs.northwestern.edu)
Holger Schultheis (schulth@sfbtr8.uni-bremen.de)
Cognitive Systems Group, Universität Bremen

Abstract

We present a computational model of mental rotation and shape comparison. The model posits that spatial abstraction, in which spatial details are removed to simplify a representation, is a key skill underlying spatial ability. Shapes are represented as collections of two-dimensional parts, and abstraction is applied by merging parts or ignoring certain part features. Using the model, we simulate a classic mental rotation experiment, demonstrating how abstraction explains the study's key finding. Finally, we compare the part-based approach to a previous edge-based approach, demonstrating that the current approach better explains human shape comparisons.

Keywords: Spatial Cognition, Cognitive Modeling, Mental Rotation.

Introduction

Spatial ability is a critical component of cognition. Children who exhibit higher spatial ability, independent of math or verbal ability, are more likely to engage in the STEM disciplines (Science, Technology, Engineering, and Mathematics) as students and later as professionals (Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009). Spatial ability is evaluated using tasks like paper-folding and mental rotation, where children must imagine a shape being transformed through space. If we can better understand the skills used on these tasks, it may be possible to teach these skills, improving students' spatial abilities and preparing them to become scientists and engineers.

We believe *spatial abstraction* is an important underlying spatial skill. Spatial abstraction is the ability to remove unnecessary detail from a mental representation, producing a simpler representation that supports faster and more effective spatial reasoning. For example, consider the leftmost shape in Figure 1A. This shape contains 12 sides. However, one might represent it as three squares stacked on top of each other, or one might even group the squares together to form one large approximate triangle.

Spatial abstraction is critical in mental rotation tasks (Figure 1A), where participants are asked whether one shape could be rotated in space to produce the other. Researchers believe individuals do this by incrementally transforming their mental representation of one shape to line it up with the other (Shepard & Metzler, 1971; Shepard & Cooper, 1982). Furthermore, there is evidence that representations are transformed in a piecemeal manner, rotating one part at a time (Yuille & Steiger, 1982; Just & Carpenter, 1976). If this is true, then the speed and ease of mental rotation will depend on the complexity of the representations. If one abstracts out details to produce a simpler representation, one can mentally rotate more effectively.

Of course, it is important to only remove the spatial detail that is unnecessary for performing a task. In mental rotation, one must keep the details that distinguish a base shape from its distractors, those items that are not valid rotations of it. Thus spatial abstraction must be sensitive to the nature of the task and the specific stimuli being processed.

Computational models can play a key role in testing abstraction strategies, concretely evaluating which details can be removed for a given task. Our previous model represented the edges going along a shape's contour, such as the 12 edges in the Figure 1A shape (Lovett & Forbus, 2013). The model encoded features for each edge and relations between edges. It performed abstraction by either a) removing the shorter edges from its representation, or b) removing all features of a particular type, e.g., ignoring the orientations of the edges.

Edge-based representation and abstraction proved effective in an initial simulation. However, this approach requires considering a great many details, such as the 12 sides in Figure 1A or the 24 sides in Figure 1B (lower row).

Here we present a new model that represents the *parts* of

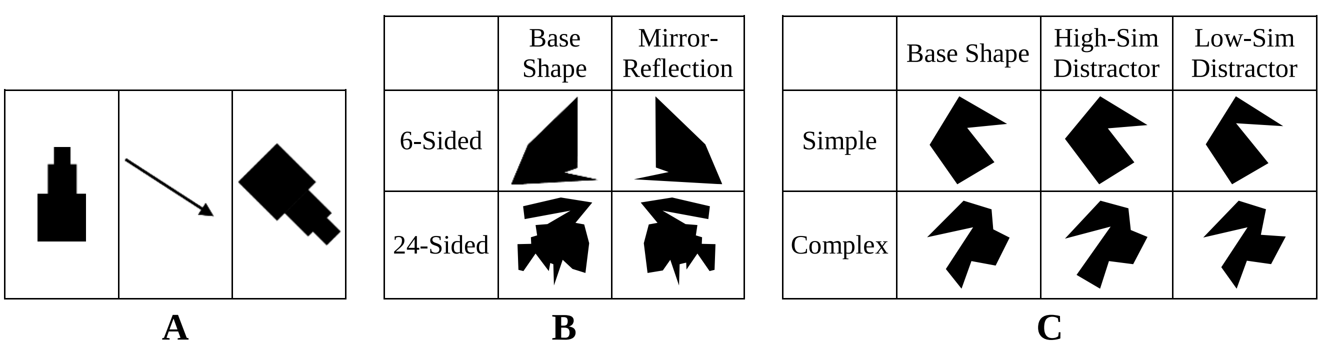


Figure 1: **A**: Mental rotation task. **B**: Stimuli from (Cooper & Podgorny, 1976). **C**: Stimuli from (Folk & Luce, 1987).

a shape, such as the stacked squares in Figure 1A. We test the part-based model on a classic mental rotation study (Folk & Luce, 1987). The model demonstrates that differing amounts of abstraction are possible, depending on the distractors. As we show, this addresses a longstanding debate in the mental rotation literature.

We begin with some background on mental rotation and spatial abstraction. We then present the motivation for a part-based approach, followed by the new model. Next we describe a simulation using two stimulus sets from Folk & Luce (1987). Finally, we compare the edge-based and part-based models, showing that the part-based approach best explains the human results.

Background

Mental Rotation

The mental rotation paradigm can provide key insights into human spatial reasoning. Consider Figure 1A, an example of sequential presentation (Cooper & Podgorny, 1976). Participants are shown the leftmost shape. Then, they are shown an arrow and instructed to mentally align the shape with the arrow's orientation. Participants must press a button when they are ready to proceed. Finally, they are shown the rightmost shape and asked whether it matches their imagined shape. In this way, researchers can isolate the time required to rotate a shape from the time required to compare shapes.

Two key questions researchers have asked are: 1) What happens to rotation time as the angle of rotation increases? 2) What happens to rotation time as the complexity of the rotated shape increases? The first question sheds light on how we represent space. If the rotation time increases at a linear rate with rotation angle, that suggests there is a mental space analogous to the physical space, and that our representation incrementally rotates through this mental space, just as an object might rotate in the physical space.

The second question sheds light on how representations move through mental space. If rotation time is constant regardless of complexity, that suggests people perform *holistic* rotation: they rotate an entire shape all at once. If rotation time increases with complexity, that suggests people perform *piecemeal* rotation: they rotate one part at a time, with more complex shapes having more parts to rotate.

There is strong, consistent evidence that as the angle of rotation increases, the rotation time increases proportionally (Shepard & Cooper, 1982). This supports the argument for a mental space analogous to physical space. However, the evidence on shape complexity is less clear, with support for both piecemeal (Yuille & Steiger, 1982; Just & Carpenter, 1976) and holistic (Cooper & Podgorny, 1976) rotation.

In one classic study, Folk and Luce (1987) demonstrated that rotation of complex shapes depended on the distractors, the shapes that were not valid rotations. Folk and Luce used six base shapes, three simple 6-sided shapes and three complex 10-sided shapes (see Figure 1C for examples). They created distractors by randomly permuting a single point on each base shape. Based on human similarity ratings, they selected a set of high-similarity and low-similarity distractors for each base shape.

Folk and Luce conducted a sequential mental rotation experiment with a blocked design. Each block contained only high-similarity or only low-similarity distractors. While no feedback was given, participants appear to have adapted to each block. In the high-similarity blocks, they rotated more slowly. More interesting was the effect of complexity. In the high-similarity blocks they rotated complex shapes more slowly than simple shapes (consistent with piecemeal rotation). However, in the low-similarity blocks there was no significant effect for complexity (consistent with holistic rotation).

Folk and Luce's results are consistent with a spatial abstraction hypothesis. Participants remove unnecessary detail to simplify their representations and speed the mental rotation. When distractors are dissimilar to base shapes, fewer spatial details are required to distinguish them, and more abstraction can be performed. This abstraction smooths out the differences between simple and complex shapes, producing a similar representation for all stimuli.

However, when distractors are similar to base shapes, more spatial details are required, and less abstraction can be performed. In this case, representations will be sensitive to the complexity of the shapes, and more complex shapes will be rotated more slowly, in a piecemeal manner.

For an abstraction process to achieve this effect, it must meet these criteria: 1) The process must be variable, able to remove more or less detail, i.e., there should be both *high abstraction* and *low abstraction*. 2) High abstraction should produce sparse representations for both simple and complex shapes, whereas low abstraction should produce denser representations, particularly dense for complex shapes. 3) One should be able to distinguish dissimilar shape pairs even after high abstraction. 4) One should be possible to distinguish *similar* shape pairs only after low abstraction.

We next consider how abstraction may be performed.

Spatial Abstraction

There are at least two types of spatial abstraction: *featural abstraction*, which removes details about the elements in a representation, and *structural abstraction*, which removes or combines the elements themselves.

Featural Abstraction Schultheis & Barkowsky (2011) argue that people use *scalable representations* to reason about space. These representations are adapted to task demands such that only the spatial information required for the task is represented. For example, in considering different object locations, one might only represent the distances between objects. However, if a task required it, one might also represent the directions between objects, e.g., noting that one is north of another. Only the necessary features will be encoded.

Consider how featural abstraction applies to an edge-based shape representation, for example describing the 12 edges in the Figure 1A shape. Each edge possesses several features, such as its location, its orientation, and its size. However, for some tasks it may be possible to abstract out certain featural dimensions—one might represent the edges as simply 12 locations in space, or 12 orientations.

Structural Abstraction Several researchers have argued that spatial representations are hierarchical, consisting of elements that can be grouped into larger structures or split into smaller pieces (Palmer, 1977; Marr & Nishihara, 1978; Lovett & Forbus, 2011). These grouping and splitting operations may be used to control the number of elements, and thus the degree of detail, in a representation.

For example, one can simply ignore some elements, believing them to be less important. In representing the 12 edges of the Figure 1A shape, one could represent just the longer edges, ignoring the shorter ones. Alternatively, one could group some elements together, producing new elements that are larger but less precise.

Part-Based Representation

Both piecewise rotation and spatial abstraction presuppose that a shape is represented as a collection of elements, with features for each elements and possibly relations between them. However, they do not specify what these elements must be. Our previous model used the edges going along a shape's contour (Lovett & Forbus, 2013). Here, we present a new approach based on the *parts* making up a shape, e.g., the three squares in Figure 1A. In line with findings on human shape processing (Hullemann et al., 2000; Hoffman & Richards, 1984), concavities of the overall shape are used to segment it into parts.

Part-based representation may support abstraction more effectively than edge-based representation for two reasons:

1) There are usually fewer parts than edges in a shape because parts are segmented at concavities, whereas edges are segmented at both convex and concave corners (e.g., three parts vs. 12 edges in Figure 1A). Thus, parts provide a more abstract starting point. There are fewer details to be considered and potentially abstracted.

2) Part-based representations strongly support structural abstraction. In any given shape, some concavities will be sharper and more salient than others. By ignoring the less sharp concavities, one can group parts together, producing a smaller set of parts that are still meaningful for representing the shape. For example, the middle shape in Figure 2 has three apparent parts. However, the lower concavity is less sharp and could be ignored, producing a simpler representation with two parts.

Model

This work builds on a model described in (Lovett & Forbus, 2011, 2013). Here we summarize the existing, edge-based model and then present the new part-based approach.

Existing Model

Representation Spatial representations can be characterized as hybrid, containing two components (Kosslyn et al., 1989): 1) A *quantitative* or metric component which describes locations, orientations, sizes, etc. These values exist in a mental space analogous to the physical space, and they can be mentally transformed. 2) A *qualitative* or propositional component, which applies labels to elements and describes how they relate to each other. For example, it might indicate that an edge is **straight** or **curved**, that one edge is **above** another, or that a corner is **concave**.

Our model builds on the CogSketch sketch understanding system (Forbus et al., 2011), which automatically generates representations from line drawings. CogSketch takes as input a set of shapes, each designated by the points going along its contour. Given a shape, it segments the contour into edges based on discontinuities in the curvature. It produces two representations to describe the edges:

1) The quantitative representation contains three values for each edge: its location, orientation, and size.

2) The qualitative representation describes the relations between the edges in a propositional form.

Abstraction The model performs featural abstraction by ignoring the locations, orientations, or sizes of the edges. It performs structural abstraction by selecting the four longest edges and ignoring all others.

Transformation The model rotates a representation by updating the three quantitative values for each edge. Any values that have already been abstracted are not transformed. After the transformation is complete, the qualitative representation is recomputed from the new quantitative values.

Comparison Before two shapes can be compared, one must determine which parts in one shape go with which parts in the other. This can be done via *structure-mapping*, a general comparison process that appears to play a role in many cognitive domains (Gentner, 1983). This process compares two propositional representations by aligning their common relational structure.

Our model compares shapes in the following manner:

1) The qualitative representations for two shapes are aligned using the Structure-Mapping Engine (Falkenhainer, Forbus, & Gentner, 1989), a computational model of structure-mapping. This identifies the corresponding edges.

2) Each corresponding pair of edges is compared using the three quantitative values. A quantitative threshold is used to determine if the values are the same. If all values are the same for all edges, the model returns a "same" response. Otherwise, it returns a "different" response.

New Model

Segmentation The part-based segmentation in the new model is based on *internal concavities* (Fairfield, 1983a). Consider Figure 2. Suppose you place a point at every location that is equidistant from two points along a shape's contour. These points make up a *Blum transform* (Blum, 1967), also known as a skeleton. Now, as you trace along the skeleton, the contour will grow closer or more distant, as the shape becomes thinner or thicker. The place where the contour stops growing closer and starts growing more distant is an inner concavity.

Fairfield (1983a) argues that inner concavities successfully capture the part segmentations that humans naturally make. The model implements the algorithm in (Fairfield, 1983b). This algorithm not only segments a shape but produces a score for each cut, based on the sharpness of the internal concavity. By default, the model allows a shape to have up to three parts. That is, it selects the two highest-score cuts and uses those to partition the shape.

Representation As with edges, the part-based representation contains quantitative and qualitative components. Because parts are two-dimensional, a fourth quantitative value has been added to the representation. Aside from location, orientation, and size (area), a part's aspect ratio is represented to better capture its shape.

Abstraction The model performs featural abstraction by ignoring certain quantitative dimensions. It performs structural abstraction by only selecting the highest-scoring cut when partitioning a shape, producing at most two parts.

Transformation & Comparison These steps are performed exactly as described for edge-based representations.

Simulation

We conducted a simulation of the Folk and Luce (1987) mental rotation experiment. As described above, this experiment used six base shapes, three simple (6-sided) and three complex (10-sided). For each base shape, there was a set of high-similarity and low-similarity distractors (see Figure 1C for examples).

Recall our criteria for an abstraction process: Low-similarity shape pairs should be distinguishable after high abstraction, which produces sparse representations for both simple and complex shapes. High-similarity pairs should be distinguishable only after low abstraction, which produces denser representations, particularly for complex shapes. Here we use the model to test several abstraction strategies and determine if they meet these criteria.

Methodology

Stimuli The simulation used nine comparators for each base shape: an identical shape, four high-similarity distractors, and four low-similarity distractors. While the original experiment included more distractors, these were the only surviving stimuli.

Each comparator was oriented at an angle 120° from the base shape. Because all rotations are mathematically equivalent for the model, it was not necessary to try each of the different rotation angles used in the experiment. In the previous simulation (Lovett & Forbus, 2013), the rotation angle was automatically computed from an arrow similar to the cue that participants would see (Figure 1A). To simplify the stimuli, the arrow was left out of this simulation, and the 120° rotation was hard-coded into the model.

Abstraction Strategies The model was run 32 times while different abstraction strategies were attempted. For featural abstraction, each of the four quantitative dimensions (location, orientation, size, aspect ratio) was used half the time and ignored half the time. For structural abstraction, the model allowed up to three parts half the time, but only up to two parts the other half.

Note that the three simple shapes only had two possible parts, while the three complex shapes had at least three (e.g., Figure 1C). Thus, when only two parts were allowed, the model produced representations of equal size for the simple and complex shapes.

For each abstraction strategy, three values were produced:

1) The fraction (0-1.0) of correct "same" responses when a base shape was compared to an identical shape.

2) The fraction of incorrect "same" responses when a base shape was compared to a high-similarity distractor.

3) The fraction of incorrect "same" responses when a base shape was compared to a low-similarity distractor.

Finally, we ran the model one additional time allowing no partitions at all. This meant that each shape was represented as a single part. This is a check on the efficacy of part-based representation, since if performance remains high, then there is no reason to segment into parts.

Results

The fraction of "same" responses for identical shapes is 1.0 across all strategies. This indicates that the model always knows when two shapes were the same. But how does the model perform when they are different?

Figure 3A shows the results when three parts are used, when two parts are allowed, or when only a single part is allowed. When three parts are used, the model is perfect at distinguishing low-similarity pairs, and near-perfect at distinguishing high-similarity pairs (<.05 error rate). When two parts are used, performance drops a small amount for low-similarity pairs but twice as much for high-similarity

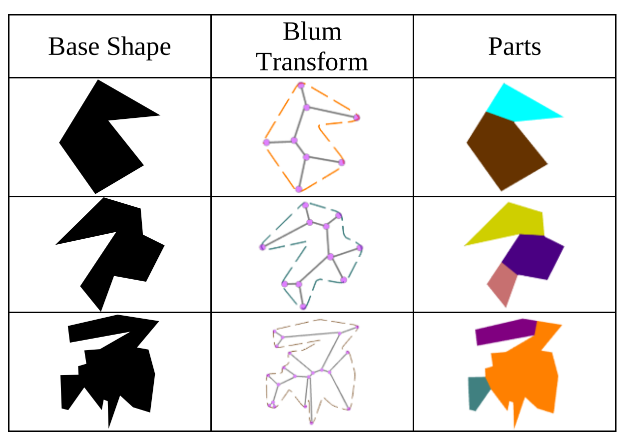


Figure 2: Part-based segmentation.

pairs. This matches the abstraction criteria, as more abstraction appears possible for the low-similarity pairs.

Finally, when only one part is used, there is a large drop in performance. Furthermore, the error rate with low-similarity pairs is actually *higher* than with high-similarity pairs, in violation of the abstraction criteria. This suggests multiple parts are needed to support effective abstraction.

We evaluated featural abstraction by beginning with the full set of four features and progressively removing the feature that was contributing the least. Figure 3B displays the results, with darker bars for when three parts are allowed, and lighter bars for when only two parts are used.

Consider first the results for low-similarity (blue). As location and orientation are removed, there is no drop in performance. These features are not needed to distinguish the shapes—only size and aspect ratio are helpful. Up to this point, the model performs perfectly with three parts, and at < .05 errors with two parts. However, when size is removed, there is a large drop in performance, suggesting that both size and aspect ratio contribute meaningfully.

Now consider the results for high-similarity (red). Again, there is no cost for removing location. However, every other abstraction is costly, raising the error rate above .1.

Discussion

The simulation results support our claims: 1) Part-based representations are effective for comparing shapes. 2) More spatial abstraction is possible when comparing low-similarity shape pairs than when comparing high-similarity shape pairs. If we select a threshold of < .05 errors, the model can achieve this using size + aspect ratio and only two parts on the low-similarity pairs, but it requires orientation + size + aspect ratio and three parts on the high-similarity pairs. This matches the criteria for high and low abstraction described above. Because high abstraction uses only two parts for both simple and complex shapes, it helps explain why people mentally rotated these at the same rate when comparing dissimilar shapes.

Interestingly, the model indicated that only size and aspect ratio were useful in distinguishing the low-similarity shape pairs. These two features would not actually need to be transformed: as an object rotates through space, they remain constant. However, we know participants were mentally rotating, as their response times increased with rotation angle. So, one might ask, what was being rotated?

We speculate that participants represented and rotated part locations, even though locations were not a necessary feature for the task. There are at least two reasons they may have done this: 1) Because participants were told to imagine the shape rotating, they needed to rotate *something*, so it would be natural to encode at least one transformable feature. 2) Participants may use locations to line up the parts when comparing two shapes, e.g., aligning the top part with the top part, even though the model demonstrated that location was not strictly necessary.

Comparison with Previous Model

We conducted two comparisons between the current model and the previous edge-based model: we ran the current model on the previous stimuli from (Cooper & Podgorny, 1976), and we ran the previous model on the current stimuli.

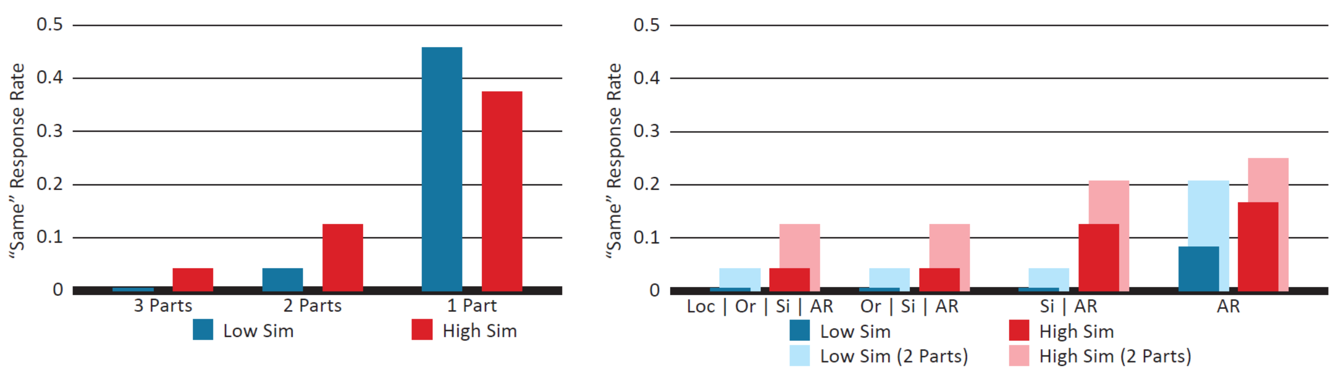
Comparison on Cooper & Podgorny Stimuli

Cooper and Podgorny used a similar sequential mental rotation paradigm. However, their five base shapes varied in complexity from 6-sided to 24-sided (see Figure 1B). They used a variety of distractors, but the previous simulation considered only one type: mirror reflections, generated by reflecting a shape over the vertical axis.

Previously, the edge-based model was able to distinguish the base shapes from the mirror reflections with 0 errors, even when only four edges were considered and when only location or only orientation was encoded for each edge.

The part-based model also makes 0 errors, even when only two parts are allowed and when only location or orientation is encoded for each part. Overall, it uses half as many features (two parts vs. four edges).

The high accuracy of both models may not be surprising, as mirror reflections are relatively easy to distinguish. However, it is notable that the key features here (orientation and location) are entirely different from the key features in the new simulation (size and aspect ratio). This highlights the point that participants must be sensitive to the task and



A **B**

Figure 3: Incorrect “same” responses with structural abstraction (A) and featural/structural abstraction (B).
Key: Loc = Location | Or = Orientation | Si = Size | AR = Aspect Ratio

the distractors when selecting features for abstraction.

Comparison on Current Stimuli

When the edge-based model is run without abstraction, it makes 0 errors distinguishing both the low-similarity and high-similarity shape pairs. However, when the model performs structural abstraction, removing all but the four longest edges, it violates the abstraction criteria: It makes no errors on the harder high-similarity pairs, but it shows a .25 error rate on the easier low-similarity pairs.

We believe this finding results from a flaw in the four-longest-edges heuristic. The assumption is that the longest edges are the most salient, and thus the most likely to be encoded. However, if participants are encoding parts, rather than edges, then a particular edge's salience may be irrelevant. It may be that changes to *shorter* edges are more noticeable because they have a greater effect on the parts (e.g., on a part's aspect ratio). Thus, an abstraction strategy that ignores shorter edges will lose valuable information.

Conclusions

Part-based representations appear to be a powerful tool in mental rotation and comparison. They allow shapes to be represented compactly but precisely. They also support spatial abstraction, in which parts are merged or features are removed. However, as our simulations show, spatial abstraction cannot be used indiscriminately. It requires careful consideration of the distractors in a task, to determine which spatial details are important.

In the future, we plan to study the process of evaluating spatial details and determining which are important. Because this process is necessary for effective abstraction, we believe it may lie at the heart of spatial ability. If we can understand how individuals isolate and exploit the key spatial details for a task, then we can teach students to be better spatial thinkers.

Acknowledgments

This research is supported by a junior fellowship from Hanse-Wissenschaftskolleg in Delmenhorst, Germany.

References

Blum, H. (1967). A transformation for extracting new descriptors of shape. In W. Wathen-Dunn (Ed.), *Models for the Perception of Speech and Visual Form* (pp. 362-380). Cambridge, MA: MIT Press.

Cooper, L., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception & Performance, 2*(4), 503-514.

Fairfield, J. (1983a). Segmenting blobs into subregions. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*(4), 363-384.

Fairfield, J. (1983b). Segmenting dot patterns by voronoi diagram concavity. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5*(1), 104-110.

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The structure mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1-63.

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception & Performance*, *13*(3), 395-404.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch understanding for cognitive science research and for education. *TopiCS*, *3*(4), 648-666.

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

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *8*(1–3), 65–96.

Hulleman, J., te Winkel, W., Boselie, F. (2000). Concavities as basic features in visual search: Evidence from search asymmetries. *Perception & Psychophysics*, *62*(1), 162-174.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441-480.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: H P & P, 15*, 723-735.

Lovett, A., & Forbus, K. (2013). Modeling spatial ability in mental rotation and paper-folding. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*.

Lovett, A., & Forbus, K. (2011). Cultural commonalities and differences in spatial problem-solving: A computational analysis. *Cognition*, *121*(2), 281-287.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Philosophical Transactions of the Royal Society of London B, 200*(1140), 269-294.

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

Schultheis, H., & Barkowsky, T. (2011). Casimir: An architecture for mental spatial knowledge processing. *Topics in Cognitive Science*, *3*(4), 778-795.

Shea, D.L., Lubinski, D., & Benbow, C.P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*, 604–614.

Shepard, R. S., & Cooper, L. A. (1982). *Mental Images and their Transformations*. Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.

Wai, J., Lubinski, D., & Benbow, C. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835.

Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics, 31*(3), 201-209.