Spatial Routines for Sketches:
A Framework for Modeling Spatial Problem Solving

Andrew Lovett, Northwestern University

Visual problem-solving is an effective tool for evaluating people's cognitive abilities. Tests such as geometric analogy and Raven's Progressive Matrices (Figure 1) ask individuals to identify commonalities and differences across images, and to transform images, applying spatial transformations to generate novel shapes and arrangements. These tasks require intelligent visual encoding (Dehaene et al., 2006), correspondence-finding across images (Primi, 2001), spatial visualization ability (McGee, 1979; Bethell-Fox, Lohman, & Snow, 1984), and a robust working memory capacity (Carpenter, Just, & Shell, 1990). Furthermore, the high correlation between performance on these tasks—particularly Raven's Matrices—and performance on other non-visual ability tests suggests that they tap into a person's general cognitive ability (Snow, Kyllonen, & Marshalek, 1984; Snow & Lohman, 1989; Burke & Bingham, 1969; Zagar, Arbit, & Friedland, 1980).

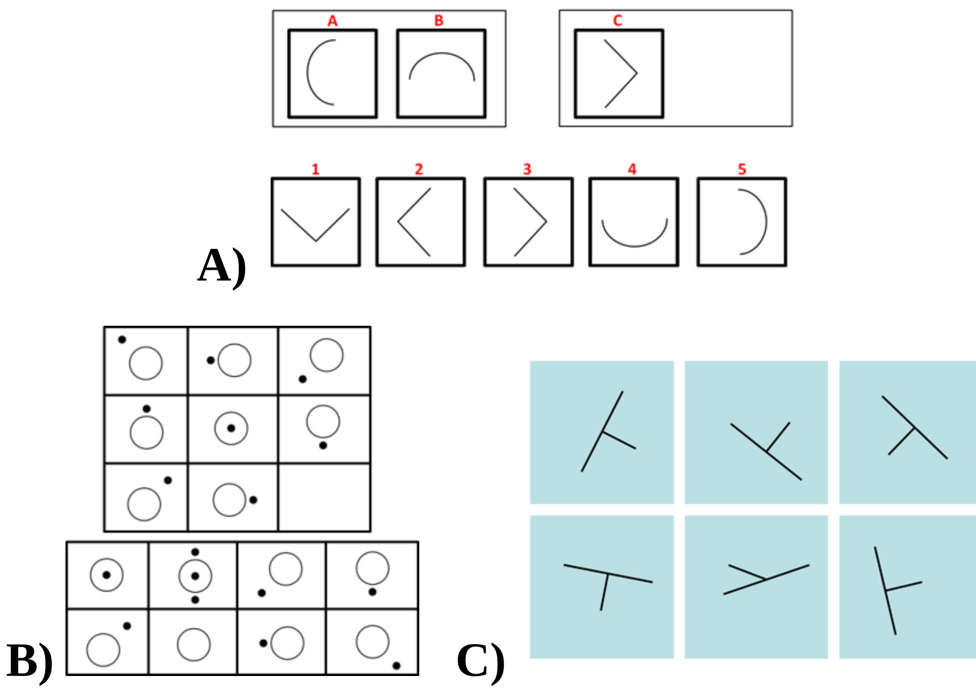

Figure 1. Three spatial problem-solving tasks.

A: Geometric Analogy: "A is to B as C is to…?" (Evans, 1968; Lovett et al., 2009b)

B: Raven's Progressive Matrices: "Pick the image that best completes the matrix." (not an actual test problem)

C: Oddity Task: "Pick the image that doesn't belong." (Deheane et al., 2006)

For my thesis, I developed a general computational framework for modeling human visual problem-solving. This framework serves two purposes: to evaluate theories of visual perception, visual comparison, and spatial visualization, the process of transforming one's visual representations (McGee, 1979); and to explore general cognitive ability, as it is evaluated in these tests. The framework builds on three psychological claims about visual processing:

1) People typically reason about space using *qualitative* or *categorical* representations (Kosslyn et al., 1989; Huttenlocher, Hedges, & Duncan, 1991; Forbus, Nielsen, & Faltings, 1991). These might indicate that one object is **right of** another or that one edge is **longer than** another, abstracting out the exact locations and sizes to facilitate comparison and reasoning. Such representations can be modeled symbolically, using predicate calculus. However, people can also use *quantitative* representations when necessary. This means the models of spatial reasoning must be able to access numerical information as well.

2) Spatial representations are hierarchical, describing an image at different levels of abstraction (Palmer, 1977; Marr & Nishihara, 1978; Navon, 1977; Hochstein & Ahissar, 2002). For example, Figure 2 might be represented as three triangular groups, nine individual objects, or 36 edges within the objects. Qualitative and quantitative information can be accessed at each of these levels. The spatial hierarchy is critical in visual problem-solving because one step in the problem-solving process is identifying the level at which a problem can be solved. Thus, flexibly moving between levels of abstraction is an important problem-solving skill and may be a core cognitive ability.

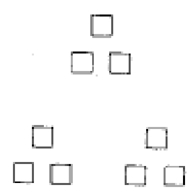

Figure 2. Stimulus from Navon (1977).

3) Qualitative spatial representations can be compared via structure-mapping (Gentner, 1983, 1989). Structure-mapping is a domain-general process in which two representations are compared by aligning their common relational structure. It was first proposed to explain abstract analogies, but there is considerable evidence that its principles also govern concrete visual comparison (Markman & Gentner, 1996; Lovett et al., 2009a; Lovett & Forbus, 2011). Structure-mapping may play a ubiquitous role in visual problem-solving because it a) determines the corresponding elements in two representations; b) highlights commonalities; and c) identifies differences.

Within my modeling framework, I built three task models: geometric analogy, Raven's Matrices, and the oddity task. A key strength of these models is that they automatically generate spatial representations from two-dimensional line drawings (input as sequences of points). Thus, they avoid the classic problem of hand-coded inputs that may be overly tailored to the model by the experimenters.

The models were evaluated by comparing their performance to that of humans on a common set of problems. All three models performed as well as typical humans, and problems that were hard for the models were also hard for people. These results support the underlying psychological claims. Furthermore, the models provided novel insights about human visual problem-solving. By selectively removing a model's ability to perform certain operations and identifying the problems it could no longer solve, I generated hypotheses about which cognitive operations are easier or harder for people to perform.

As described below, these hypotheses address questions about the cultural variability of spatial concepts and factors affecting reasoning more broadly.

Modeling Framework

The modeling framework, *Spatial Routines for Sketches*, builds on Ullman's (1984) visual routines proposal. Ullman suggested that visual processing could be divided into a set of *basic operations*, such as tracing along a curve or exploring the region within a curve. These operations could be parameterized and combined to create a *visual routine*, a strategy for computing some visual property from an image.

A visual routine is analogous to a computer program, with each operation acting as a procedure. Thus, the visual routine proposal lays out a clear strategy for building computer models of visual processing: 1) Identify a set of basic operations, cognitive operations that people can perform during processing. 2) Implement each operation as a procedure in a computational language. 3) Write routines in that language. Each routine is both a cognitive model, describing the strategy someone might use to perform a task, and a computer program.

Spatial Routines for Sketches applies the visual routine concept to high-level visual problem-solving, rather than low-level visual processing. It relies on three types of operations: 1) *visual perception*, which generates a qualitative, symbolic representation from an image; 2) *visual comparison*, which compares two representations to identify commonalities and differences; and 3) *visual inference*, which applies differences to one image to infer a new image (e.g., the answer in geometric analogy problems). These operations can be parameterized and combined to create a *spatial routine*, a model for solving a task such as geometric analogy.

The operations are summarized below:

1) *Visual perception* generates spatial representations from two-dimensional line drawings. The operation models human perception at Marr's computational level. That is, its purpose is to create human-like representations, not to model the processes that produce those representations. It can be parameterized to generate representations at three hierarchical levels: individual edges, objects, or groups of objects. Its representations are qualitative and symbolic. However, the symbols point back to actual edges, objects, and groups in the image. When necessary, the system can return to these elements to query for quantitative information (see shape comparison below).

Visual perception is integrated with the CogSketch sketch understanding system (Forbus et al., 2011). Modelers can use CogSketch to create stimuli, either sketching out shapes by hand or importing from another program such as PowerPoint. All of the stimuli in my thesis (e.g., Figure 1) were imported from PowerPoint, eliminating the need for hand-coding representations.

2) *Visual comparison* identifies commonalities and differences in two representations. According to structure-mapping theory, people compare symbolic representations by aligning the common relational structure (Gentner, 1983, 1989). This process highlights commonalities but also draws attention to differences tied to the common structure. Structure-mapping is implemented using the Structure-Mapping Engine (SME) (Falkenhainer, et al 1989), a well-established computational model.

Spatial Routines always uses SME for comparison, but it implements two operations that differ in their outputs. *Generalize* produces a symbolic representation of the common elements it identifies. For example, comparing the first two images in Figure 1B would produce a generalization with a large circle and a small circle, the large circle lying **right of** the small circle. In contrast, *Find-Differences* produces a representation of the changes between two or more images. Here, it would describe a qualitative change where one object ceases to be **above** the other.

Because structure-mapping is a domain-general comparison process, it can operate on any symbolic representation, not only image representations. Thus, it is possible to generalize across sets of differences, or to find the differences between generalizations. This flexible use of comparison is the key to effective problem-solving.

Sometimes visual problem-solving requires comparing shapes. In Figure 1A, one must determine that a shape is rotated clockwise 90°. This requires moving beyond the qualitative, symbolic representation to compute a quantitative shape transformation. The operation computes shape transformations through a three-step process, based upon my extension of the psychological research on mental rotation (Shepard & Metzler, 1971; Shepard & Cooper, 1982): 1) Use structure-mapping to find the corresponding edges in the two shapes. 2) Compute a quantitative transformation (e.g., a rotation) over one pair of corresponding edges. 3) Apply this transformation to the full set of edges in one shape, evaluating whether the transformed edges quantitatively match the edges in the other shape. This model's use of structure-mapping over qualitative representations allows it to explain why people typically know which direction to mentally rotate one shape to align it with another: they are guided by the initially pair of corresponding parts.

3) *Visual inference* applies differences to one image to produce a novel image representation. It models spatial visualization (McGee, 1979), that is, the ability to manipulate one's spatial representations. At the symbolic level, it adds or removes qualitative relations, for example, inferring in Figure 1B that the missing image in the lower right should contain a small circle **right of** and **below** the large one. It also applies quantitative shape transformations, e.g., rotating the shape in Figure 1A.

Task Models

I built three task models within the Spatial Routines framework: geometric analogy, Raven's Matrices, and the oddity task. The three models use identical initial parameters for perception, meaning they start by computing visual representations in exactly the same way. They begin with group-level representations (recall the groups of squares in Figure 2), based on the hypothesis that people generally begin with high-level visual representations and work down the hierarchy as needed to perform a task (Hochstein & Ahissar, 2002; Navon, 1977).

Each task model was compared to human performance on a common set of problems. The evaluation asked three questions: 1) Can the model perform at human level, that is, does it solve as many problems as typical humans? 2) Do the model's error patterns match human error patterns? 3) Can the model provide novel insights about human problem-solving? The final question was evaluated as follows:

1) Code each problem according to which particular model operations are required to solve it. For example, some problems may only be solved if the model uses a certain level in spatial hierarchy. The coded-for operations varied by task.

2) Code each problem for working memory load—the number of objects the model considers when solving the problem.

3) Conduct a linear regression to determine how well these factors explain human response times or error rates.

In this way, the models can suggest which cognitive operations are particularly difficult for people, as well as evaluated the role of working memory load in problem difficulty.

Geometric Analogy (Lovett et al., 2009b, Lovett & Forbus, 2012)

In the geometric analogy task (Figure 1A), test-takers are asked, "A is to B as C is to…?" Geometric analogy is a classic problem-solving task, and it has been modeled many times (e.g., Evans, 1968; O'Donoghue, Bohan, & Keane, 2006; Schwering et al., 2009). However, the present model is the first to integrate two competing theories for how people perform it. The first strategy (Sternberg, 1977), which I term the *visual inference strategy*, works as follows:

1) Compare images A and B to find the differences.

2) Compare images A and C to find the corresponding elements.

3) Apply the A/B differences to the corresponding elements in C to infer the answer image. Compare the inferred answer to each listed answer.

The second strategy, which I term the *second-order comparison strategy* (Evans, 1968; Mulholland, Pellegrino, & Glaser, 1980), works as follows:

1) Compare images A and B to find the differences.

2) For each possible answer *i*, compare C to that answer to get the differences.

3) Compare the A/B differences to the C/*i* differences. Pick the answer *i* that produces the most similar differences.

Whereas previous models have followed one strategy or the other, the present model implements both. The model is able to integrate these strategies because structure-mapping is a general-purpose comparison process: it can compare either images, or the abstract differences *between* images. Because the visual inference strategy is more efficient, requiring fewer comparisons, the model first attempts this strategy. When it is unable to infer an answer image, or when the inferred image fails to match any listed answer, it reverts to second-order comparison, reusing the results of the initial A/B comparison (step 1 above). Thus, the model makes concrete predictions about when people will take longer to solve a problem.

The model was evaluated on 20 problems from Evans (1968). Human participants were given the same problems, and their answers and response times were measured. Overall, the model chose the answer preferred by humans on all 20 problems. Because of this task's relatively low difficulty, further analysis was restricted to human response times—error rates were not considered. The linear regression included the number of objects, whether the model reverted to second-order comparison, and whether the model made additional strategy shifts during problem-solving. The regression found that these factors together accounted for .95 of the variance in human response times (i.e., $R^2 = .95$). Participants took longer on problems with more objects and on problems where the model predicted there would be strategy shifts.

Importantly, the best match to humans was achieved when the working memory load measure was the number of objects in the A/B differences, not the number of objects in the answer image. This suggests people have more difficulty keeping abstract differences in memory, compared with concrete image representations. It raises the intriguing possibility that spatial working memory limits may depend on the content being remembered. Due to our familiarity with concrete objects, they may be processed more fluidly and efficiently than abstract differences.

Raven's Matrices (Lovett, Forbus, & Usher, 2010; Lovett & Forbus, in prep)

Raven's Progressive Matrices (e.g., Figure 1B) is a visual intelligence test. Test-takers are shown a (typically) 3x3 matrix of images with the lower-right image missing and asked to solve for it. The Spatial Routines model of Raven's Matrices, uses the same strategies as the geometric analogy model, providing a test of the approach's generality. Given a problem, it first attempts the visual inference strategy, applying the differences in the upper rows to the bottom row to infer the answer image. If it fails, it reverts to second-order comparison, inserting each possible answer into the bottom row and comparing the resulting row representation to the above rows.

Importantly, the model also uses the same representations as the geometric analogy model: it calls the *visual perception* operation using the exact same parameters. By evaluating a single representation scheme across multiple tasks, I can better determine how well it captures human spatial reasoning.

Raven's Matrix problems are designed to challenge people's intelligence, often requiring backtracking and strategy shifts to solve. Thus, this model has several built-in strategies for comparing images, and for representing the differences across a row. Note that while other Raven's models have also included built-in strategies (Carpenter, Just, & Shell, 1990; Rasmussen & Eliasmith, 2011), the present model's strategies are composed within the Spatial Routines framework and tied to specific psychological processes.

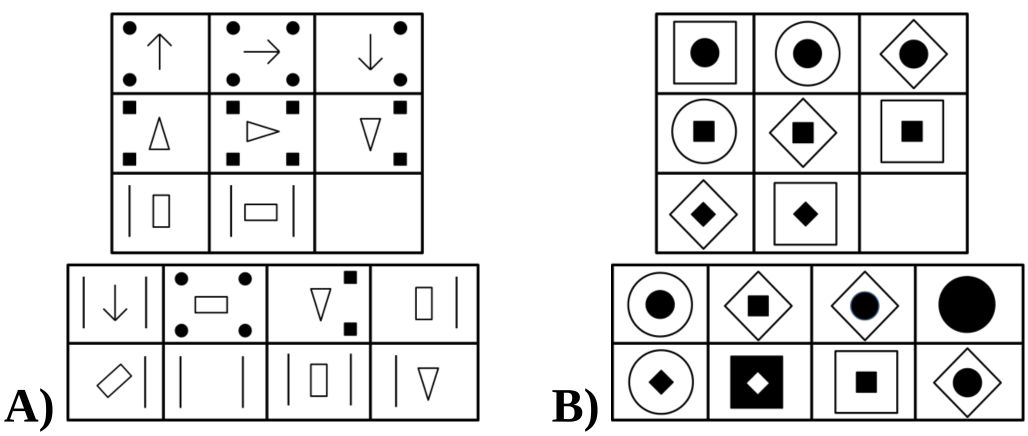

Figure 3. Raven's Matrix problems requiring particular strategies to solve.

One key process in visual problem-solving is *perceptual reorganization*. This is when an image is re-represented and divided into a different set of meaningful elements to support comparison and problem-solving. For example, consider Figure 3A (to protect the security of the test, no actual Raven' problems are shown here—these examples were designed by the author). In the initial representations, each top-row image contains two elements: an arrow, and a group of circles. When the first two images are compared via structure-mapping, the arrow maps to the arrow, and the group maps to the group. At this point, the model moves down the hierarchy to compare the parts of the corresponding elements. It compares the edges in the two arrows, determining that there is a 90° rotation between them. It compares the circles in the two groups, determining that the circles in the left image map to the leftmost circles in the central image. Similarly, it determines that rightmost circles in the central image map to the circles in the right image. Based on these correspondences, the model manipulates its own spatial representation. It divides the circles in the central image into two separate groups: the left column and the right column. This helps the model develop a more meaningful representation of the row: one shape rotates clockwise, while an element is added to its right and then removed from its left.

While perceptual reorganization is helpful, and sometimes critical, for visual problem-solving, it may be difficult to perform: it requires flexibly moving between levels in the spatial hierarchy while keeping previous comparison results in memory. Thus it is one of the factors I considered when analyzing human performance on this task.

In addition to re-representing an image, it may be necessary to re-represent an entire row. Typically (e.g., Figure 3A), rows are represented as differences between adjacent images, similar to the A/B differences in geometric analogy. However, in Figure 3B the important feature is not the *differences* across a row, but the common elements in each row. Here, each row contains a square, a circle, and a diamond, but their order varies across rows. To handle such problems, the model implements two strategies for representing a row: a *differences* strategy and a *common-elements* strategy. The model selects a strategy by comparing the top two rows and evaluating which strategy makes the rows most similar.

The Raven's Matrix model was evaluated on the Standard Progressive Matrices, a 60-problem intelligence test (Raven, Raven, & Court, 1989). The model included a low-level visual processing strategy for dealing with simpler problems that lack the 3x3 matrix. Overall, the model solved 56/60 problems, placing it in the 75th percentile for American adults according to 1993 norms.

To further evaluate the model, I built a computerized version of the test to gather human performance data. The computerized test covered only the 36 problems containing 3x3 matrices, that is, the hardest problems on the test. 42 Northwestern University undergraduates took this test. Data from two participants was discarded because their scores fell more than two standard deviations below the mean. Among the remaining 40 students, the mean score was 30.0/36, compared to the model's 32/36 on the same problems.

The model's four failed problems were among the five hardest for the human participants. Thus, problems that were hard for the model were also hard for people. Further analysis was conducted on 31 problems, removing the four failed problems and one additional problem, which people find difficult for reasons outside the model's scope.

The 31 problems were coded for number of elements and for which strategy shifts were required to solve them. This was done via ablation, selectively removing the ability to perform an operation and identifying the problems which could no longer be solved. A linear regression found that five factors explained .8 of the variance in human accuracy ($R^2 = .8$). There was a significant cost for perceptual reorganization.

Interestingly, the analysis found a cost for number of elements (working memory load) specifically for problems involving a *differences* strategy, rather than a *common-elements* strategy. As with geometric analogy, working memory load was a factor when participants were remembering abstract differences between images, rather than image contents.

These results, taken with the geometric analogy results above, suggest what makes someone an effective visual problem-solver. Skilled visual problem-solving depends on a robust working memory, and particularly an ability to remember and process abstract representations, such as differences between images. Furthermore, it requires an ability to reorganize one's spatial representations, in order to construct the necessary representations for solving a problem.

Note that a memory for abstract details and the ability to reorganize representations are not only useful in the visual domain. These may be two of the general abilities that tie visual problem-solving to other cognitive domains. Further research by psychologists and modelers is required to evaluate these abilities and their broader role in human reasoning.

The Oddity Task (Lovett & Forbus, 2011)

Deheane et al. (2006) designed a visual oddity task (Figure 1C) to investigate spatial understanding in different cultures. Participants are shown six images and asked to choose the one that doesn't belong. Each problem requires understanding of a particular spatial concept to solve it. For example, Figure 1C requires understanding perpendicular lines. Other problems involve orientation, concavity, containment, alignment, symmetry, and rotation.

Dehaene and colleagues administered a 45-problem oddity task to both North Americans and the Mundurukú, a South American indigenous group. They found that the Mundurukú performed above average on most problems, despite no formal schooling in geometric concepts, or exposure to words for most of the concepts. They took this as evidence for a universal geometric module, innately available to all humans. However, there is reason to believe *some* aspects of the task are learned. The North Americans did better as they got older, whereas Mundurukú of all ages performed the same on the task (Newcombe & Uttal, 2006).

I developed an oddity task model to explore how spatial reasoning varies across cultures. The model builds on the hypothesis that while certain *processes* for spatial reasoning are universal, cultural groups may vary in their fluency with different spatial *representations*. Specifically, I hypothesize that structure-mapping across qualitative representations is a universal mechanism for visual comparison and visual problem-solving. Groups may vary in how well they encode certain spatial features and how well they utilize different levels in the spatial hierarchy.

The model solves problems by generalizing over a subset of the answers (e.g., the top or bottom row), and then comparing each remaining image to the generalization. If one image is noticeably less similar, that image is chosen as the odd one out.

The model computes qualitative image representations in exactly the same way as the previous models. It begins at the top of spatial hierarchy, with group-level representations (e.g., the three groups of squares in Figure 2). If it fails to find an image that is noticeably less similar, it moves down the hierarchy. Thus, the model makes two universal predictions: 1) People will identify the odd image out more easily if it varies *qualitatively* from the other images. 2) In order to solve a problem, people will need to identify the particular level in the spatial hierarchy where the odd image out becomes salient.

I evaluated the model using the 45 problems from (Dehaene et al., 2006), comparing it to four human groups: young children (North Americans, age 4-8), older children (North Americans, age 8-12), adults (North Americans, age 18-52), and the Mundurukú (age 5-83). The Mundurukú were not divided into age groups because their performance did not vary by age.

The model correctly solved 39/45 problems, or 87%. This is comparable to North American adults, the highest human performers, who were at 83%. Furthermore, the model's error patterns (assigning each problem 1.0 for correct or 0 for incorrect) correlated with human accuracy values on each problem. The highest correlation was with American adults ($r = .77$), while the lowest was with the Mundurukú ($r = .49$).

Further analysis was restricted to the 39 solved problems. As before, each problem was coded for which operations were needed to solve it, as well as for number of elements. A linear regression was run for each group's accuracy on the 39 problems, and the results were compared. There were both cultural commonalities and differences. All groups had trouble with problems involving shape comparison. To solve these problems, one must first compare shapes within a single image and encode a relationship (for example, the image might contain a *rotation* between shapes). One must then compare *across* images to identify the image that lacks this relationship. Such problems may be difficult because shape comparison itself is difficult (Bethell-Fox, Lohman, & Snow, 1984) or because they require that participants spend extra time encoding the internal structure of a single image before comparing images.

Other problems were difficult for one group but not for the other. The pattern of results suggests the groups differed in their fluency with each level of the spatial hierarchy. The North Americans had trouble with problems requiring edge-level representations (this factor was significant for North American children and marginally significant for adults). In contrast, the Mundurukú had trouble with problems requiring group-level representations. This suggests a cultural difference in visual processing and spatial ability. North Americans can easily see large-scale objects and object groupings. However, they find it difficult to ignore the overall object and focus on the parts within it. In contrast, the Mundurukú can easily see the parts within an object but have difficulty grouping objects together into a larger scene. This may be because of education differences, e.g., North Americans are trained in basic geometric shapes and can easily see the patterns that groups of shapes make. Additional research is needed to evaluate this novel psychological hypothesis.

In this task, the number of elements (working memory load) was not a significant factor. This is a consistent finding. In the previous tasks, the number of elements was a factor specifically in abstract representations of image differences. This task did not require processing abstract differences—instead of representing what is changing between images, the model merely identifies the elements that are common across all images and then finds an image that lacks those elements. Thus, this result further supports the claim that abstract differences place a greater load on working memory.

Conclusion

Spatial Routines for Sketches is an effective framework for modeling visual problem-solving. It allows multiple tasks to be modeled using a common set of operations. The models provide evidence for the framework's psychological claims. They demonstrate that visual problem-solving can be performed using structure-mapping across qualitative, hierarchical representations. The models' error patterns match human error patterns, providing further support.

The models also generate novel psychological hypotheses. They suggest that skilled visual problem-solving depends on 1) a robust working memory capacity, particularly for abstract representations, 2) the ability to reorganize one's representations to better fit a problem, and 3) an inclination to explore and encode an item's internal structure (e.g., to compare the shapes within a single image of the oddity task). Because these are general cognitive abilities that also apply to non-visual domains, they help explain why these tasks are effective for evaluating general intelligence.

Moreover, the oddity task model suggests that different cultural groups may use the same cognitive processes but vary in their fluency with each level of the spatial hierarchy. Whereas previous researchers have argued that people universally start with high-level representations and work their way down (Hochstein & Ahissar, 2002; Navon, 1977), the Mundurukú appear to find *low-level* representations more accessible, and to have difficulty using high-level representations (groups of objects like the squares in Figure 2).

This thesis is heavily interdisciplinary. It uses computational models both to evaluate psychological theories and to generate cognitive hypotheses. New psychological and neuroscientific studies will be required to test these hypotheses. In the future, such studies can help to refine the models and enhance our understanding of human cognition.

References

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*, 205-238.

Burke, H. R., & Bingham, W. C. (1969). Raven's Progressive Matrices: More on construct validity. *Journal of Psychology: Interdisciplinary and Applied, 72*(2), 247-251.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review, 97*(3), 404-431.

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

Evans, T. (1968). A program for the solution of geometric-analogy intelligence test questions. In M. Minsky (Ed.), *Semantic Information Processing*. Cambridge, MA: MIT Press.

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

Forbus, K., Nielsen, P., and Faltings, B. (1991). Qualitative spatial reasoning: The CLOCK project. *Artificial Intelligence, 51*(1-3), 417-471.

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

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

Gentner, D. (1989). Mechanisms of analogical learning. In S. Vosniadou and A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 199-241). London: Cambridge University Press.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791-804.

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

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: Human Perception and Performance, 15*, 723-735.

Lovett, A., & Forbus, K. (in prep). Modeling comparison and abstraction during visual problem-solving.

Lovett, A., & Forbus, K. (2012). Modeling multiple strategies for solving geometric analogy problems. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Sapporo, Japan.

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

Lovett, A., Forbus, K., & Usher, J. (2010). A structure-mapping model of Raven's Progressive Matrices. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Lovett, A., Gentner, D., Forbus, K., & Sagi, E. (2009a). Using analogical mapping to simulate time-course phenomena in perceptual similarity. *Cognitive Systems Research 10*(3): Special Issue on Analogies - Integrating Cognitive Abilities, 216-228.

Lovett, A., Tomai, E., Forbus, K., & Usher, J. (2009b). Solving geometric analogy problems through two-stage analogical mapping. *Cognitive Science, 33*(7), 1192-1231.

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

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, Series B, Biological Seicnece, 200*(1140), 269-294.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*(5), 889-918.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12*, 252-284.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

Newcombe, N. C., & Uttal, D. H. (2006). Whorf versus Socrates, Round 10. *Trends in Cognitive Science, 10*(9), 394-396.

O'Donoghue, D. P., Bohan, A., & Keane, M. T. (2006). Seeing things - Inventive reasoning with geometric analogies and topographic maps. *New Generation Computing, 24*(3), 267-288.

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

Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence, 30*, 41-70.

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science, 3*(1), 140-153.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford: Oxford Psychologists Press.

Schwering, A., Gust, H., Kühnberger, K., & Krumnack, U. (2009). Solving geometric proportional analogies with the analogy model HDTP. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge: MIT Press.

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

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of learning and ability correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (vol. 2, pp. 47–103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational Measurement, 3rd ed.* (pp. 263–331). New York, NY: Macmillan.

Sternberg, R. J. (1977). *Intelligence, Information Processing, and Analogical Reasoning*. Hillsdale, NJ: Erlbaum.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-159.

Zagar, R., Arbit, J. & Friedland, J. (1980). Structure of a psychodiagnostic test battery for children. *Journal of Clinical Psychology, 36*, 313 - 318.