

Modeling Cross-Cultural Performance on the Visual Oddity Task

Andrew Lovett¹, Kate Lockwood¹, and Kenneth Forbus¹

¹ Qualitative Reasoning Group, Northwestern University
2133 Sheridan Rd., Evanston, IL, 60201, USA
{andrew-lovett@, kate@cs., forbus@}northwestern.edu

Abstract. Cognitive simulation offers a means of more closely examining the reasons for behavior found in psychological studies. This paper describes a computational model of the visual oddity task, in which individuals are shown six images and asked to pick the one that doesn't belong. We show that the model can match performance by participants from two cultures: Americans and the Mundurukú. We use ablation experiments on the model to provide evidence as to what factors might help explain differences in performance by the members of the two cultures.

Keywords: Qualitative representation, analogy, cognitive modeling, oddity task

1 Introduction

A central problem in studying spatial cognition is representation. Understanding and modeling the visual representations people construct for the world around them is a difficult challenge for cognitive science. Dehaene and colleagues [7] made important progress on this problem by designing a study which directly tests what features people represent when they look at geometric figures in a visual scene. Their study utilized the *Oddity Task* methodology: participants were shown an array of six images and asked to pick the image that did not belong (e.g., see Fig. 1). By varying the diagnostic spatial feature, i.e., the feature that distinguished one image from the other five, they were able to test which features their participants were capable of representing and comparing.

Dehaene and colleagues ran their study on multiple age groups within two populations: Americans and the Mundurukú, an indigenous group in South America. They found that while the Americans performed better overall, the Mundurukú appeared to be capable of encoding the same spatial features. The Mundurukú performed above chance on nearly all of the 45 problems, and their pattern of errors correlated highly with the American pattern of errors. Dehaene concluded from the results that many spatial features are universal in human representation. However, several questions remain: (1) What makes one problem harder than another? (2) Why is it that, despite the high correlation between population groups, some problems seem especially hard for Americans, while other problems seem especially hard for the Mundurukú? (3) To what extent can questions 1) and 2) be answered in terms of

the process of encoding representations, versus the process of operating over those representations to solve problems?

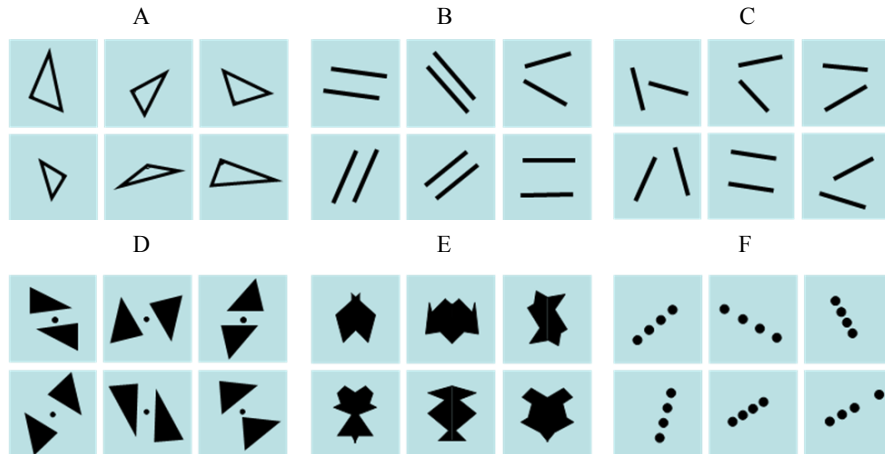


Fig. 1. Six example problems from the Oddity Task.

This paper presents a cognitive model designed to explore these questions. Our model is based upon two core claims about spatial cognition: (1) When people encode a visual scene, they focus on the *qualitative* attributes and relations of the objects in this scene [11]. This provides them with a more abstract, more robust representation than one filled with quantitative details about the scene. (2) People compare low-level visual representations using the same mapping process used to perform abstract analogies. Our model of comparison is based on Gentner’s [14] structure-mapping theory of analogy.

Our model uses four components to simulate the oddity task from end-to-end. We use a modified version of CogSketch¹ [13], a sketch understanding system, to automatically construct qualitative representations of sketches and other two-dimensional stimuli. We use the Structure-Mapping Engine (SME) [8], a computational model of structure-mapping theory, to model comparison and similarity judgments. We use two additional components based on structure-mapping theory: MAGI [9], which models symmetry detection, and SEQL [18], which models analogical generalization. Using this approach, we have modeled human performance on geometric analogy problems [25] (problems of the form “A is to B as C is to ...?”); a subset of the Raven’s Progressive Matrices [20], a visually-based intelligence test; and basic visual comparison tasks [19,21]. However, the Dehaene task offers a unique opportunity in that it was designed to isolate specific spatial features and check for their presence or absence in one’s representation.

This paper presents our cognitive model of performance on the Oddity Task and uses it to study factors that contribute to difficulty on the task. In comparing the model with human results, we focus on two population groups: American children

¹ Publicly available at http://www.spatialintelligence.org/projects/cogsketch_index.html

aged 8-13, and the full set of Mundurkú of all ages. We consider these groups because their overall performance on the 45 problems in the Dehaene study was comparable: 75% for the Americans and 67% for the Mundurkú. We provide evidence for what might distinguish these groups from each other via ablation studies using the model.

We begin by briefly reviewing the components of our model. We then show how these component models are combined in our overall model of the Oddity Task. We analyze the results produced by running the model on the 45 problems from the original study, and use ablation studies to explore possible explanations for performance differences between the two groups. We close with a discussion of related and future work.

2 Modeling Comparison via Analogy

Our model of comparison is based on Gentner's [14] structure-mapping theory of analogy. According to structure mapping, people compare two objects by aligning the common structure in their representations of the objects. Comparison is guided by a *systematicity* bias; that is, people prefer mappings that place deeper structure with more higher-order relations into correspondence. Structure-mapping has been used to explain and predict a variety of psychological phenomena, including visual similarity and differences [15,22]. Next we describe three computational models based on structure-mapping, each of which is used in the present study.

2.1 Structure-Mapping Engine

The Structure-Mapping Engine (SME) [8,10], is a computational implementation of structure-mapping theory. It takes as input two cases, a *base* and a *target*. Each case is a symbolic representation consisting of *entities*, *attributes* of entities, and *relations*. There are both first-order relations between entities and higher-order relations between other relations. SME returns one to three mappings between the base and target. Each mapping has three components: a set of *correspondences* between elements in the base and elements in the target, a *structural evaluation score*, which estimates the degree of similarity between the cases; and a set of *candidate inferences*, inferences about the target supported by the mapping and unaligned structure in the base.

2.2 MAGI

MAGI [9] is a model of symmetry detection based upon SME. Essentially, it identifies symmetry in a representation by comparing the representation to itself, while avoiding perfect self-matches. MAGI is important in modeling spatial cognition because it is often necessary to identify axes of symmetry in a visual scene, or in a specific object.

2.3 SEQL

SEQL [18] is a model of analogical generalization. SEQL is based upon the idea that individuals learn generalizations for categories through a process of *progressive alignment* [16], in which instances of a category are compared and the commonalities are abstracted out as a direct result of the comparison. Given a set of cases, SEQL can build one or more generalizations from them by comparing them via SME and eliminating the structure that fails to align between cases, leaving only the structure that is common across all the cases in the generalization. Because the generalization is in the same form as individual case representations, new cases can be compared to the generalization to measure their similarity to a category.

3 Modeling Qualitative Representation via CogSketch

One of our core claims is that people use qualitative spatial representations when reasoning over or comparing images. While quantitative data, such as the exact sizes of objects or the exact orientation of edges, may vary widely, even between images of the same object, qualitative relations are much more consistent. For example, nearly every face contains an eye to the right of another eye, with both eyes above a nose and a mouth. The key to qualitative representation is to encode what Biederman calls the *nonaccidental properties* [4]. These are the relations that are unlikely to have occurred by accident in a sketch. For example, two lines chosen at random are unlikely to have exactly the same orientation. Therefore, when two lines are parallel, this is unlikely to have occurred by random chance, and so it is probably significant.

There is abundant evidence that people encode qualitative relations corresponding to nonaccidental properties in visual scenes. For example, parallel lines are salient for children as young as three [1]. Adults and infants can distinguish between concave and convex shapes—a qualitative distinction [3], and humans have been shown to have a preference for objects aligned with a vertical or horizontal axis, as opposed to those with an arbitrary orientation [2]. Huttenlocher and colleagues [17] have shown that when individuals memorize a point’s location in a circle, they pay special attention to which quadrant of the circle the point lies in, again a qualitative distinction. While it is obviously the case that individuals are capable of encoding quantitative information in addition to these qualitative relations, the qualitative relations appear particularly well-suited to spatial problem-solving, as they can be easily encoded symbolically and used to compare different scenes. Thus, in our present work we explore the hypothesis that spatial tasks can be solved relying exclusively on qualitative representations.

We see qualitative spatial representations as hierarchical (e.g., [24]). Each of the shapes in an image can have its own attributes and relations. At the same time, each of the edges that make up that shape can also have its own attributes and relations. This gives rise to two representational foci: the *shape representation* and the *edge representation*. A further claim we are evaluating with the current study is that these two representational foci will never be used together. That is, a comparison or other

operation will always run on either an image’s shape representation or its edge representation.

3.1 CogSketch

CogSketch [13] is a sketch understanding system based upon the nuSketch [12] architecture. Users sketch a series of *glyphs*, or objects in a sketch. CogSketch then computes a number of qualitative spatial relations between the glyphs, building up a structural representation of the sketch that corresponds to the *shape* representation. CogSketch can also decompose a glyph into its component edges and construct a representation of the qualitative relations between the glyph’s edges. This corresponds to the *edge* representation.

Many of the spatial relations in the shape representation (e.g., relative position, containment) are computed based on the relative position and topology of the glyphs. However, some shape relations can only be computed by first decomposing a glyph into its edges and constructing the glyph’s edge representation. By comparing two glyph’s edge representations using SME, CogSketch can identify the corresponding edges in the two glyph’s shapes. These correspondences can be used to determine whether the two glyphs are the same shape, and whether one glyph’s shape is a transformation of the other (e.g., a rotation or a reflection). Furthermore, a glyph’s edge representation can be compared to itself via MAGI to identify axes of symmetry.

Table 1. Qualitative vocabulary for the edge representation.

<u>Edge Attributes</u>	<u>Edge Relations</u>
<ul style="list-style-type: none"> • Straight/Curved/Ellipse • Axis-aligned (horizontal or vertical) • Short/Med/Long (relative length) 	<ul style="list-style-type: none"> • Concave/convex corner • Perpendicular corner • Edges-same-length corner • Intersecting • Parallel • Perpendicular

3.2 Representing the Oddity Task stimuli

In order to model the Oddity Task, we examined the Dehaene [7] stimuli and identified a set of qualitative attributes and relations that appeared to be important for solving the problems. All attributes and relations had to be among those that could be computed automatically by CogSketch.

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

Table 2. Qualitative vocabulary for the shape representation.

<u>Shape Attributes</u> <ul style="list-style-type: none"> • Closed shape • Convex shape • Circle shape • Empty/Filled • Axis (Symmetric, Vertical, and/or Horizontal) 	<u>Shape Relations</u> <ul style="list-style-type: none"> • Right-of/Above (relative position) • Containment • Frame-of-Reference • Shape-proximity-group • Same-shape • Rotation-between • Reflection-between
<u>Line-Line Relations</u> <ul style="list-style-type: none"> • Intersecting • Parallel • Perpendicular 	<u>Line-Point Relations</u> <ul style="list-style-type: none"> • Intersecting • Colinear • Centered-On

Table 2 summarizes attributes and relations for shapes. *Empty/filled* is a simplification of shape color; it refers to whether the shape has any fill color. *Frame-of-Reference* relations are used when a smaller shape is located inside a larger, symmetric shape (i.e., a circle). The inner's shape location is described in terms of which quadrant of the larger shape it is located in; additionally, the inner shape may lie exactly along the larger shape's axes of symmetry. *Shape-proximity-group* refers to shapes grouped together based on the Gestalt law of proximity [26]. Currently, grouping by proximity is only implemented for circles.

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

Axes of symmetry, *same-shape*, *rotation-between*, and *reflection-between* are all computed by comparing shapes' edge representations, as described above. Reflections are classified as *X-Axis-Reflections*, *Y-Axis-Reflections*, and *Other-Reflections*.

4 Modeling the Oddity Task

Our approach to performing the Oddity Task is to identify what is common across the images in an array by generalizing over their representations with SQL. Individual images can then be compared to the generalization using SME. If one image is noticeably less similar to the generalization, then it must be the odd image out. Most of the time (e.g., Problem B in Fig. 2), the odd image out lacks a qualitative feature that is present in the other five images, in this case parallel lines. However, in some cases (e.g., Problem C), the odd image out possesses an extra feature beyond those found in the other images.

4.1 Theoretical Claims of Model

Our model of the oddity task is based on the following theoretical claims:

- 1) People encode qualitative, structural representations of visual scenes and use these representations to perform visual tasks.
- 2) For a given problem, people will focus on a particular representational level (either the shape level or the edge level) in solving that problem.
- 3) Qualitative spatial representations are compared via structure-mapping, as implemented in SME.
- 4) People will identify the common features across a set of images via analogical generalization, as implemented in SEQL.

Note that these claims are general enough to apply to many spatial tasks. However, they are not detailed enough to fully specify how any task would be completed. Thus, it is necessary to make additional modeling assumptions in order to fill out a complete computational model of the task.

4.2 Modeling the Process

Our model attempts to pick out the image that does not belong by performing a series of Generalize/Compare trials. In each trial, the system constructs a generalization from a subset of the images in the array (either the top three or the bottom three). This generalization represents what is common across all of these images. For example, consider the right-angled triangle problem (Fig. 1, Problem A). The generalization built from the three top images will describe three connected edges, with two of the edges being perpendicular. In the rightmost top image, the two perpendicular edges form an *edges-same-length-corner*, but this relation will have been abstracted out because it is not common to all three images.

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

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

Alternatively, similarity scores can be normalized based on both the size of the generalization and the size of the image's representation (*fully-normalized*). This score measures both how much of the generalization is present in the image and how much of the image is present in the generalization. While more complex than *gen-normalized* scores, *fully-normalized* scores are necessary for noticing an odd image out that possesses an extra qualitative feature that the other images lack. For

example, it allows the model to pick out the image with parallel lines from the other five images without parallel lines (Fig. 1, Problem C).

4.3 Controlling the Processing

In each Generalize/Compare trial, the model must make three choices. The first is which subset of the images to generalize over (either the top three images or the bottom three). The second is whether to use gen-normalized or fully-normalized similarity scores. The third is whether to use edge representations or shape representations—recall that we are predicting that edge representations and shape representations will never be combined in a single comparison.

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

After the initial pair of trials is run, the model looks for a *sufficient* candidate. Recall that each Generalize/Compare run produces three similarity scores for the three images that have been compared to the generalization. A sufficient candidate is chosen when the lowest-scoring image has a similarity score noticeably lower than the other two ($< 95\%$ of the second lowest-scoring image), meaning the image is noticeably less similar to the generalization.

In cases where a sufficient candidate is not found, the model will attempt additional trials. (1) If the model was previously run using edge representations, it will try using shape representations. (2) The model will try using a fully-normalized similarity score, to see if the odd image out possesses an extra feature. At this point, if no sufficient candidate has been identified, the model gives up (this is the equivalent of a person guessing randomly, but we do not allow the model to make such guesses).

5 Simulation

We evaluated our model by running it on the 45 problems from the Dehaene [7] study. The original stimuli, in the form of PowerPoint slides, were copied and pasted into CogSketch, which automatically converted each PowerPoint shape into a glyph. Four of the 45 problems were touched up in PowerPoint to ease the transition—lines or polygons that had been drawn as separate parts and then grouped together were redrawn as a single shape. Five additional problems were modified after being pasted

into CogSketch. In all five cases, we removed simple edges which had been added to the images of the problem to help illustrate an angle or reflection to which participants were meant to attend. Because the model was unable to understand the information these lines were meant to convey, they would have served only as distracters. Aside from the changes to these nine problems, no changes were made to the stimuli which had been run on human participants.

In analyzing the results, we consider first the model's overall accuracy, including the correlation between its performance and that of both the American participants and the Mundurukú participants. We then use the model to identify four factors that could contribute to problem difficulty. We examine the correlation between these factors and human performance on the subset of problems that are correctly solved by the model.

5.1 Model Accuracy

Our model correctly solves 39/45 problems. Note that chance performance would be 7.5/45. Furthermore, there is a strong correlation between the model's performance and the performance of the human participants. Table 3 shows the Pearson correlation coefficient between the model and each of the human populations. As the table shows, the model correlates better with the American participants. However, there is also a high correlation with the Mundurukú participants. The coefficient of determination, which is computed by squaring the correlation coefficient, indicates the percentage of the variance in one variable which is accounted for by another. In this case, the coefficient of determination between the model and the Mundurukú participants is $(.493^2 = .243)$, meaning the model accounts for about $\frac{1}{4}$ of the variance in the performance of the Mundurukú participants.

Table 3. Correlations between the model and the American and Mundurukú participants.

	Americans	Mundurukú
Model	.656	.493
Americans	*	.758
Mundurukú	.758	*

Fig. 2 plots the performance of the two populations and the model. As the figure shows, the six problems on which the model fails are among the hardest for both populations. The one clear exception is problem 21 (see Fig. 3). Although the model fails on this problem, the Mundurukú performed quite well on it (86% accuracy).

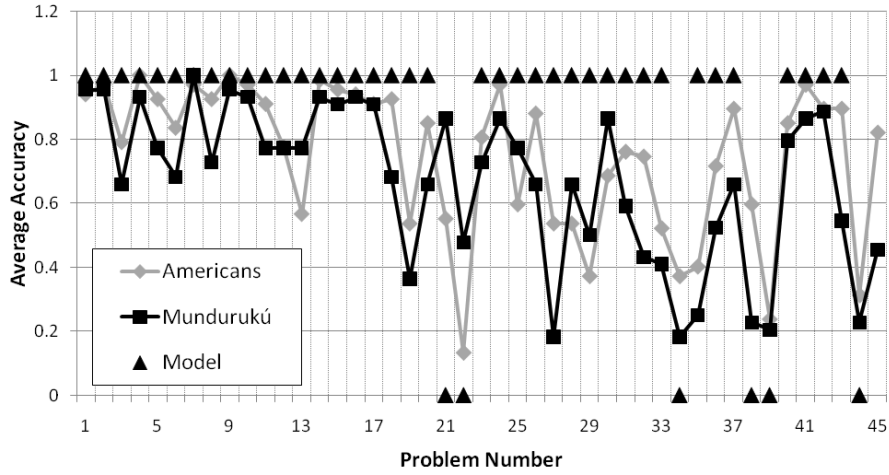


Fig. 2. Performance of Americans, Mundurukú, and our model on the Oddity Task.

Discussion. Fig. 3 shows the six problems which our model fails to solve. As the percentages show, these problems were for the most part quite difficult for both the Americans and the Mundurukú, with performance on some problems little or no higher than chance (17%).

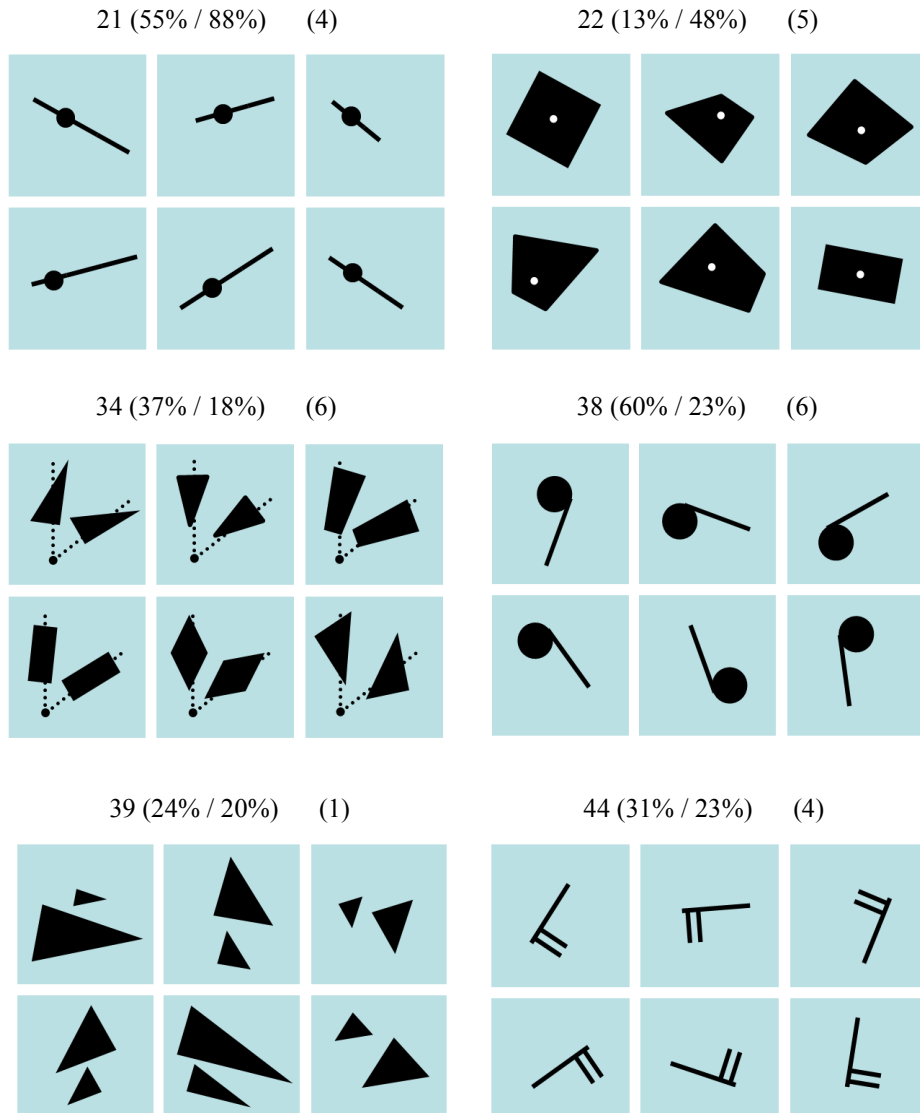


Fig. 3. The six problems the model fails to solve. Above each problem the average accuracy for the Americans and the Mundurukú are listed, respectively, followed by the number of the correct answer.

Overall, these six problems can be roughly broken down into three categories based on what is required to solve them. First, problem 22 requires encoding whether the dot lies along the axes of the asymmetric quadrilateral. Our model simply does not encode this relation—nor, it appears, do Americans, as they actually fall below chance on this problem. Interestingly, the Mundurukú are well above chance; at this time, it is difficult to say why they are better at solving this problem.

Problems 38 and 44 both require identifying a rotation between shapes found in different images. Our model only looks for rotations between shapes within a single image. As the percentages show, the participants, and particularly the Mundurukú, had difficulty solving these problems. We believe the most likely reason is that it did not occur to them to look for rotations between shapes in different images.

Problems 21, 34, and 39 all appear to require encoding a quantitative relation between shapes: a percentage distance along an edge for 21, a number of degrees of rotation for 34, and a ratio between two shapes' sizes for 39. The fact that participants had so much trouble with these problems supports our prediction that individuals primarily encode and compare *qualitative* spatial features. The one exception here was problem 21, which was reasonably difficult for the Americans but actually quite easy for the Mundurukú. As with problem 22, it is difficult to say why the Mundurukú performed so well on this problem. It may be that they are better at dividing a space (either a line or a quadrilateral) into smaller parts and qualitatively encoding which of those smaller parts a dot lies along.

5.2 Modeling Problem Difficulty

We analyzed problem difficulty on the 39 problems that the model correctly solves. We used the model to identify four factors that could contribute to difficulty. For this paper, we focus on factors related to encoding the stimuli. The factors are:

(1) *Shape Comparison*. Some problems (e.g., Fig. 1, Problem D) require constructing edge representations of two shapes and comparing them in order to identify a relation between the shapes (e.g., a rotation or a reflection). This may be difficult because it involves switching between the edge and shape representations, and because it requires conducting an additional comparison with SME before one begins comparing the six images.

(2) *Shape Symmetry*: Some problems (e.g., Fig. 1, Problem E) require comparing a shape's edge representation to itself, via MAGI, in order to identify an axis of symmetry. This could be difficult for similar reasons.

(3) *Shape Decomposition*: Several problems (e.g., Fig 1, Problem A) require decomposing shapes into edges in order to represent each image at the edge representation level. It is possible that this will be difficult for individuals because there may be a temptation to consider closed shapes only at the shape representation level.

(4) *Shape Grouping*: A couple problems (e.g., Fig. 1, Problem F) require grouping shapes together based on the Gestalt rule of proximity. Normally, one would assume this was easy, but preliminary analysis indicated it might be difficult for the Mundurukú participants.

We used the model to produce a measure for each difficulty factor on each problem via ablation; for example, we ran the model with the ability to conduct shape comparisons turned off in order to identify the problems on which shape comparisons were required. We then attempted to find a difficulty function, based on the four factors, which correlated highly with each of the human populations. This was done by performing an exhaustive search over all possible linear weights for the four factors in the range of 0 to 15.

Results. The optimal difficulty function for the American participants is shown in Table 4 (the weight for each factor is normalized based on the size of the largest weight). In addition to the weight of each factor, the table shows the individual contribution of each factor to the correlation between the function and human performance. This was computed by removing a factor from the difficulty function and considering the drop in the function's correlation with the human population.

As Table 4 shows, the difficulty function had an overall correlation of .667 with the American participants. This means that the function explains ($.667^2 = 44\%$) of the variance in human performance on the 39 problems. Most of the contribution to this correlation comes from shape comparison and shape symmetry. It appears that the American participants had a great deal of difficulty with problems that required decomposing shapes into edges and comparing the edge representations to identify relations between shapes, or symmetry within a single shape. Shape decomposition also contributed to the correlation, suggesting that the participants had some difficulty with the problems requiring focusing on the edge representations of closed shapes.

Table 4. Relative contribution of factors to our difficulty function for American performance.

Factor	Weight in Function	Contribution to Correlation
Shape Comparison	.69	.163
Shape Symmetry	1.0	.267
Shape Decomposition	.38	.062
Shape Grouping	.08	.001
Overall	---	.667

The optimal difficulty function for the Mundurukú participants is shown in Table 5. This difficulty function had a correlation of .637 with the human data, indicating it accounts for ($.637^2 = 41\%$) of the variance in the Mundurukú performance. By far, the most important factor was shape comparison. The other contributing factor was shape grouping, suggesting that the Mundurukú participants might have some difficulty with problems requiring grouping elements together based on proximity. This is surprising, as Gestalt grouping is generally thought to be a basic, low-level operation. Note that the Mundurukú participants had no trouble with problems requiring estimating relative distances, as indicated by their high performance on problem 21 (Fig. 3).

Table 5. Relative contribution of factors to our difficulty function for Mundurukú performance.

Factor	Weight in Function	Contribution to Correlation
Shape Comparison	1.0	.393
Shape Symmetry	.29	.018
Shape Decomposition	.14	.009
Shape Grouping	.71	.081
Overall	---	.637

Table 6 shows the correlation between each difficulty function and each population group. As expected, each difficulty function correlates far better with the population group for which it was built. The fact that there is still a relatively high correlation between the American function and the Mundurukú performance, and between the Mundurukú function and the American performance, most likely results from the fact that both groups have a great deal of trouble with problems requiring shape comparison.

Table 6. Correlations between difficulty function and population.

Difficulty Function	American Participants	Mundurukú Participants
American Function	.667	.402
Mundurukú Function	.427	.637

Discussion. One of our original goals was to use the model to identify differences between the two populations. Our two difficulty functions appear to have accomplished this. The difficulty function for American participants suggests that they tend to encode images holistically. They tend to have trouble when a problem requires breaking a shape down into its edge representation. This may be because the academic training in basic shapes encourages Americans to look at shapes as a whole, rather than explicitly considering the individual edges that make up a shape. The Mundurukú participants, in contrast, appear to encode stimuli more analytically. They are better able to consider shapes in terms of their component edges; most noticeably, they are better at using a shape's edges to identify axes of symmetry. However, they had difficulty seeing groups of shapes holistically in this task.

6 Related Work

Several AI systems have been constructed to explore visual analogy. Croft and Thagard's DIVA [5] uses a 3D scene graph representation from computer graphics as a model of mental imagery. That is, the system "watches" animation in the computer graphics system in order to perceive its mental imagery. Analogy is carried out via a connectionist network over the hierarchical structure of the scene graph. DIVA's initial inputs, unlike ours, are generated by hand. Their background knowledge is also hand-generated specifically for their simulation, unlike our use of the same knowledge base across many simulation systems and experiments. DIVA has only been tested on a handful of examples, and to the best of our knowledge, has not been used to model specific psychological findings. Davies and Goel's Galatea [6] uses a small vocabulary of primitive visual elements (line, circle box) plus a set of visual transformation over them (e.g., move, decompose) to describe base and target descriptions, and uses a copy/substitution algorithm to model analogy, carrying sequences of transformations from one description to the other. All of Galatea's inputs are hand-generated, as is its background knowledge, and it has only been tested on a few examples. Mitchell and Hofstadter's Copycat [23] modeled analogy as an aspect of high-level perception, using comparisons between letter strings as the domain. Copycat was domain-specific, and even the potential correspondences between items were hand-coded (the slipnet), making it less flexible than SME, which is domain-independent.

7 Discussion

We have described a model of the Oddity task, using CogSketch to automatically encode stimuli in terms of qualitative spatial representations, MAGI to detect symmetry, and SME and SEQL to carry out the task itself. We showed that this combination of modules can achieve behavior comparable to the participants in Dehaene *et al's* study of American and Mundurukú performance on the same stimuli. Furthermore, we were able to provide some evidence about possible causes for performance differences between the groups, through statistical analysis of ablation experiments on the model.

We find these results quite exciting on their own, but they are also part of a larger pattern. That is, similar combinations of qualitative representations and analogical processing have already been used to model a variety of visual processing tasks [19,20,25]. This study lends further evidence for our larger hypotheses, that (1) qualitative attributes and relations are central to human visual encoding and (2) people compare low-level visual representations using the same mapping process they use for abstract analogies. The study also lends support to the proposal that (3) comparison operations are performed using either a shape representational focus or an edge representational focus.

We plan to pursue two lines of investigation in future work. First, this paper focused on difficulties related to encoding. Our model suggests difficulties involving comparisons may also be implicated. For example, a problem might be harder

because the six images in the array are less similar, making alignment and generalization production more difficult. We plan to explore how well aspects of the comparison process can explain the variance. Of particular interest are whether their contributions are universal, or whether there will be cultural differences. Second, we plan on using these analyses to construct more detailed models of specific groups performing this task (i.e., children and adults, as well as both cultures). Comparing these models to each other, and to models of similar spatial tasks, could help identify general processing constraints on such tasks. This may shed light on how universal human spatial representations and reasoning are, both across cultures and across tasks.

Acknowledgements

This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC). We thank Elizabeth Spelke for providing the original oddity task stimuli.

References

1. Abravanel, E.: The Figure Simplicity of Parallel Lines. *Child Development*, 48(2), 708-710 (1977)
2. Appelle, S.: Perception and Discrimination as a Function of Stimulus Orientation: The "Oblique Effect" in Man and Animal. *Psychological Bulletin*, 78, 266-278 (1972)
3. Bhatt, R., Hayden, A., Reed, A., Bertin, E., Joseph, J.: Infants' Perception of Information along Object Boundaries: Concavities versus Convexities. *Experimental Child Psychology*, 94, 91-113 (2006)
4. Biederman, I.: Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, 94, 115-147 (1987)
5. Croft, D., Thagard, P.: Dynamic Imagery: A Computational Model of Motion and Visual Analogy. In: L. Magnani and N. Nersessian (eds.) *Model-based Reasoning: Science, Technology, Values*, pp. 259-274, Kluwer/Plenum, New York (2002)
6. Davies, J., Goel, A.K.: Visual Analogy in Problem Solving. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 377-382 (2001)
7. Dehaene, S., Izard, V., Pica, P., Spelke, E.: Core Knowledge of Geometry in an Amazonian Indigene Group. *Science*, 311, 381-384 (2006)
8. Falkenhainer, B., Forbus, K., Gentner, D.: The Structure-Mapping Engine. In: *Proceedings of the Fifth National Conference on Artificial Intelligence* (1986)
9. Ferguson, R.W.: MAGI: Analogy-Based Encoding Using Regularity and Symmetry. In: *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, pp. 283-288 (1994)
10. Forbus, K., Oblinger, D.: Making SME Greedy and Pragmatic. In: *Proceedings of the Cognitive Science Society* (1990)
11. Forbus, K., Ferguson R., Usher, J.: Towards a Computational Model of Sketching. In: *Proceedings of the 2001 Conference on Intelligent User Interfaces (IUI-2001)*
12. Forbus, K., Lockwood, K., Klenk, M., Tomai, E., Usher, J.: Open-Domain Sketch Understanding: The nuSketch Approach. In: *AAAI Fall Symposium on Making Pen-based Interaction Intelligent and Natural* (2004)

13. Forbus, K., Usher, J., Lovett, A., Wetzel, J.: CogSketch: Open-Domain Sketch Understanding for Cognitive Science Research and for Education. In: Proceedings of the Eurographics Workshop on Sketch-Based Interfaces and Modeling (2008)
14. Gentner, D.: Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7(2), 155-170 (1983)
15. Gentner, D., Markman, A.B.: Structure Mapping in Analogy and Similarity. *American Psychologist*, 52, 42-56 (1997)
16. Gentner, D., Loewenstein, J.: Relational Language and Relational Thought. In: Amsel, E., Byrnes, J.P. (eds.) *Language, Literacy, and Cognitive Development: The Development and Consequences of Symbolic Communication*. Lawrence Erlbaum Associates (2002)
17. Huttenlocher, J., Hedges, L.V., Duncan, S.: Categories and Particulars: Prototype Effects in Estimating Location. *Psychological Review*, 98(3), 352-376 (1991)
18. Kuehne, S., Forbus, K., Gentner, D., Quinn, B.: SEQL: Category Learning as Progressive Abstraction Using Structure Mapping. In: Proceedings of the 22nd Annual Meeting of the Cognitive Science Society (2000)
19. Lovett, A., Gentner, D., Forbus, K.: Simulating Time-Course Phenomena in Perceptual Similarity via Incremental Encoding. In: Proceedings of the 28th Annual Meeting of the Cognitive Science Society (2006)
20. Lovett, A., Forbus, K., Usher, J.: Analogy with Qualitative Spatial Representations Can Simulate Solving Raven's Progressive Matrices. In: Proceedings of the 29th Annual Conference of the Cognitive Society (2007)
21. Lovett, A., Sagi, E., Gentner, D.: Analogy as a Mechanism for Comparison. In: Proceedings of Analogies: Integrating Multiple Cognitive Abilities (2007)
22. Markman, A.B., Gentner, D.: Commonalities and Differences in Similarity Comparisons. *Memory & Cognition*, 24(2), 235-249 (1996)
23. Mitchell, M.: *Analogy-making as Perception: A Computer Model*, MIT Press, Cambridge, MA (1993)
24. Palmer, S.E.: Hierarchical Structure in Perceptual Representation. *Cognitive Psychology*, 9(4), 441-474 (1977)
25. Tomai, E., Lovett, A., Forbus, K., Usher, J.: A Structure Mapping Model for Solving Geometric Analogy Problems. In: Proceedings of the 27th Annual Conference of the Cognitive Science Society (2005)
26. Wertheimer, M.: Gestalt Theory. In: W. D. Ellis (ed.) *A Sourcebook of Gestalt Psychology*, pp. 1-11, The Humanities Press, New York (1924/1950)