Ainooson, James, and Kunda, Maithilee (2017). A computational model for reasoning about the Paper Folding task using visual mental images. In Proceedings of the 39th Annual Meeting of the Cognitive Science Society, London, UK.

A Computational Model for Reasoning About the Paper Folding Task Using Visual Mental Images

James Ainooson (james.ainooson@vanderbilt.edu)
Maithilee Kunda (mkunda@vanderbilt.edu)
Department of Electrical Engineering and Computer Science, PMB 351679
2301 Vanderbilt Place, Nashville, TN 37235-1679, USA

Abstract

The paper folding task is commonly used for the evaluation of nonverbal, spatial reasoning skills. In this paper, we present a computational model that attempts to use visual-imagery-based representations and operations to solve this task. The model was tested against all problems from the standard paper folding task and achieved a perfect score, illustrating that visual-imagery-based representations and operations are sufficiently expressive to capture at least one successful solution strategy. Although the model does not closely resemble human cognitive processing, and thus should not be considered in its current form to be a plausible psychological model of human task performance, the assumptions made and their implications for our understanding of human cognition on the paper folding task point to fruitful lines of future work towards this goal.

Keywords: artificial intelligence; cognitive assessment; paper folding; spatial skills.

Introduction

Paper folding tasks are cognitive assessment tools used in the evaluation of spatial, non-verbal reasoning skills. Visuospatial skills in general are thought to be critical to a variety of human endeavors, including scientific discovery (Miller, 1984), art (Arnheim, 1969), engineering (Ferguson, 1994), computer programming (Petre & Blackwell, 1999), mathematics (Giaquinto, 2007), education (Silverman, 2002), and even feats of memory (Foer, 2011). Visuospatial skills also seem to be areas of intact or even superior performance for certain individuals with developmental conditions such as autism (Soulieres et al., 2011; Kunda & Goel, 2011) and Prader-Willi syndrome (Verdine et al., 2008).

In research on Science, Technology, Engineering and Mathematics (STEM) education, visuospatial ability is viewed as a key contributor to math learning (National Research Council, 2009) and to pursuing degrees and careers in STEM disciplines (Wai et al., 2009). Studies suggest that visuospatial ability can improve with training (Uttal et al., 2013), and that such training can enhance math performance in children (Cheng & Mix, 2014).

Thus, there is an urgent need for effective visuospatial assessments as well as training interventions to promote learning outcomes, creative discoveries, effective design work, and more. Understanding the specific cognitive mechanisms that underlie visuospatial ability is a critical step along this path.

Of course, studying visuospatial ability purely through observations of human behavior is challenging because many of the underlying cognitive processes are not directly observable. Even neuroimaging yields only a coarse view of the specific information processing steps that take place as someone is solving a task.

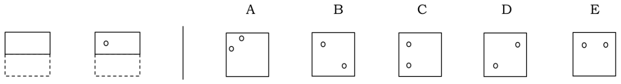

Figure 1: A sample task from the VZ-2 paper folding test. The images on the left of the vertical line depict the stages in a fold. The images on the right of the line are possible choices of how the paper may look when unfolded.

In this paper, we instead adopt the approach of implementing a computational cognitive model that simulates solving the task—the cognitive systems approach of artificial intelligence (AI) (Thagard, 2005). Cognitive systems model how intelligent agents combine different cognitive processes, like learning, reasoning, and memory, to perform a task. By implementing a cognitive system that simulates solving visuospatial tasks, we can look "under the hood" at specific types of information processing mechanisms that might drive visuospatial ability.

In previous work, we have implemented similar models that investigate aspects of visuospatial cognition on other cognitive assessment tasks. Previous work on the Raven's Progressive Matrices intelligence test has examined the role of visual mental representations in solving difficult test problems (Kunda et al., 2013), the contributions of different types of imagery operators (Kunda et al., 2013), the meaning of different patterns of errors (Kunda, Soulières, et al., 2016), and visual mechanisms for maintaining goal-subgoal hierarchies (Kunda, 2015). Previous work on the Block Design task has looked at relationships between internal mental representations and external deployments of visual attention (Kunda, El Banani, & Rehg, 2016), and previous work on the Embedded Figures task has looked at capacity limits in visuospatial memory, in particular the effects on task performance of internal deployments of visual attention to different parts of a visual mental representation (Kunda & Ting, 2015).

In this paper, we present initial results from a new computational model of the paper folding task. Although the model does not closely resemble human cognitive processing, and thus should not be considered in its current form to be a plausible psychological model of human task performance, the assumptions made and their implications for our understanding of human cognition on the paper folding task point to fruitful lines of future work towards this goal.

In particular, the model we present can be considered as an experiment on the sufficiency of certain imagery-based representations and operations for solving paper folding, which is valuable for understanding how different cognitive mechanisms might in theory contribute to visuospatial ability in people, and especially how certain cognitive limitations might affect task performance. Ultimately, we hope that results from this line of work will serve as a basis to suggest routes for how such cognitive limitations might eventually be overcome, i.e., in developing new visuospatial training interventions for use in education.

About Paper Folding Tasks

Paper folding tasks are usually presented as line-drawings of paper cut-outs or folded pieces of paper. People are then asked to imagine changes that happen when this paper is manipulated in different ways. Several forms of the test exist.

Shepard and Feng (1972) presented a form of the paper folding test which required test subjects to fold a cube out of six connected squares. Two of the squares have arrows that point to an edge and one square is greyed out to show that it is the base of the cube. People are then required to predict whether the arrows align and point to the same edge when the cube is re-constructed. This essentially requires them to mentally reconstruct the cube by imagining folding the paper as though the shape was cut-out of the paper. Lovett and Forbus (2013) developed a computational model developed to reason about this specific case of paper folding tasks. Their model takes the approach of simplifying the task by removing unnecessary details and essentially focusing primarily on the orientation of critical edges to solve the task.

Another form of the paper folding task (which is also called "the punched hole test"), developed by Ekstrom and colleagues (1976), is administered as a 6 minute pencil and paper test in two parts (3 minutes for each part). During the test, subjects are presented with a sequence of images showing the stages in folding a square piece of paper. A hole is then punched on this folded piece. Test subjects are also presented with five possible outcomes of how the paper looks when it is unfolded. See Figure 1 for an example of this type of problem. Designed as part of the "Kit of Factor-Referenced Cognitive Tests", this task appears as the second test under the group of tests that evaluate the visualisation cognitive factor (VZ-2).

A number of research studies have used paper folding to evaluate spatial reasoning skills. While testing a hypothesis on learning styles amongst individuals, Mayer and Massa (2003) used this test as part of their measure of spatial skills. Keehner et al. (2004) also used this test as one of their tests of spatial ability while investigating the correlation between spatial ability, experience and skill in laparoscopic surgery. Another example is the study by Silvia (2008) in which the paper folding test was used as part of a measure of fluid intelligence, while investigating relationships between creativity and intelligence.

There is much that is still unknown about the direct cognitive mechanisms involved in paper folding tests. Mental rotations are believed to play a major role (Shepard & Feng, 1972). In addition to the complexity of the mental rotations, people may have to deal with the additional component of mental folding (Glass et al., 2012) and in the case of a punched hole test, how the holes affect the final output. Wright and colleagues (2008) showed that training on mental rotation tasks improved performance on paper folding tasks, just as training on paper folding tasks improved mental rotations.

Next, we present a computational model that attempts to solve the paper folding task using simulated "mental rotations" in "three dimensions". The exact formulation of the paper folding task we intend to tackle with this model is "The Punched Hole" test (Ekstrom et al., 1976).

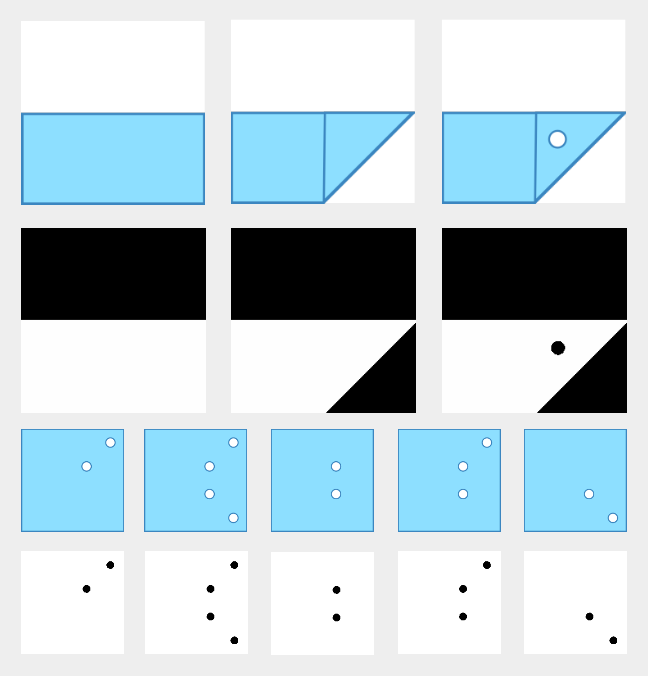

Figure 2: A sequence of images sent as input to the model (blue and white), and the corresponding bitmaps that are used by the model after inputs are processed (black and white). The first input row corresponds to the initial "problem" part of a paper folding item. The second input row contains the possible choices the model is presented with.

The Model

We present a computational model that attempts to solve "The Punched Hole" paper folding task (Ekstrom et al., 1976) using only image based operations. The model is built in the Python programming language and relies extensively on the Pillow fork of the Python Imaging Library (PIL) to perform low level image manipulation.

The main task of the model is to analyze a sequence of images that depict the folding and punching in a problem of the paper folding task to determine what the paper would look like when unfolded. It achieves this by maintaining a three dimensional representation of the paper which is stored as a stack of two dimensional images. Each image on the stack represents a single level of folding performed in a single time-slice. The actual fold operations are performed with image reflections that provide a simplified simulation of three-dimensional rotations.

Input

Inputs to the model are presented in three stages. The first stage consists of a sequence of images that represent the state of the folded paper in each time-slice. The second stage consists of a single image that represents the state of the folded paper after the hole has been punched. Finally, the third stage presents the possible solutions from which the model could select an answer after it is done predicting a solution.

All the inputs are presented as line drawings with sections that contain paper filled with pixels to ensure the model can properly differentiate between paper and empty space. Before any images are passed on for further processing, they are converted to single colour images. This makes it easy for the model to perform logical bitwise operations between images. Once converted to a single colour image, any pixel that is set to 1 in the image is considered to be an area containing paper, and pixels set to 0 are considered to be empty spaces. See Figure 2 for a sample sequence of input images and their corresponding bitmap representations.

Strategy

The model's strategy for solving the tasks relies heavily on two stacks. The first stack (which we call the *image stack*) keeps track of images that represent the layers of folds. The second stack (which we call the *operations stack*) keeps track of the operations that are performed on the images as folds occur. In predicting the solution, the model utilizes four main operations: *Initialize*, *Fold*, *Punch* and *Unfold*.

The *Initialize* operation sets up the model before solving any task. It places an image which has all its bits set to 1 on the images stack. This image is meant to represent the initial piece of unfolded paper on which fold operations will be performed.

The *Fold* operation receives as input an image of the state of the paper after a given fold has occurred. The model will attempt to use this image and other images on the image stack to find the best estimation of the line along which the fold was made. To accomplish this, the following processing steps are performed on the bitmap representation of the fold input image for every image on the stack, starting from the bottom:

1. The current layer to be processed is retrieved from the *image stack*. In the case of the first fold operation, this image has all pixels set to 1.

2. An intersection operation is performed between an inverse of the fold input image and the image retrieved from the stack. This operation is constrained by the bounding box around the image that was retrieved from the stack. The resulting image is an image of the flap to be folded, as illustrated in Figure 3.

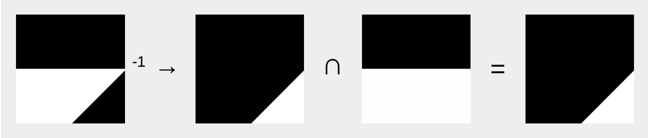

Figure 3: Stages the images go through to generate the folded flap image.

3. Another intersection operation is performed between the fold input and the image retrieved from the stack. This new image replaces the original image on the stack, as shown in Figure 4.

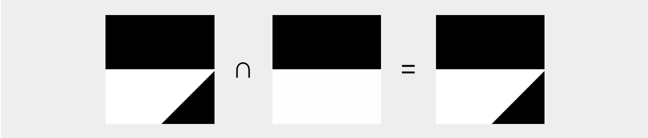

Figure 4: The intersection between the input image and the existing image on the stack to generate a replacement for the image on the stack.

4. In order to determine the line along which the fold would occur, a single pixel border is drawn around both the folded flap image, and the modified image on the stack. This is to make both images larger so they can slightly overlap. An intersection operation is computed between the two new overlapping images to generate an image containing the line along which the fold is to be made. A search for two extreme coordinates of this image is then performed on the pixels in this image. This search is biased towards the pixels that are from the folded flap image. Search results will now contain the coordinates of the fold line. These coordinates are then pushed onto the operations stack.

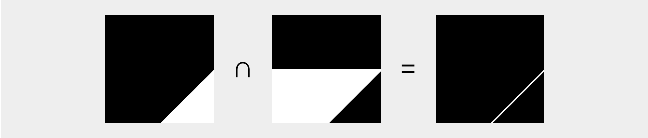

Figure 5: Intersecting the new overlapping images to determine the fold line.

5. Finally, the folded flap image is reflected across the fold line. This reflection operation is analogous to a 180° rotation in three dimensional space about the fold line axis. The reflected image is then pushed to the top of the stack to act as one of the base images for any subsequent fold operations. After a series of fold operations is performed, each image on the stack will represent a folded layer.

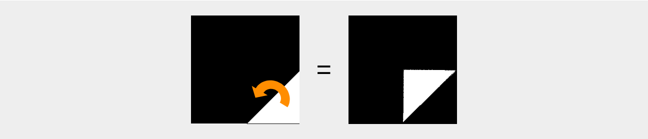

Figure 6: The final reflection operation of the fold.

Once all fold operations are completed, the *Punch* operation is performed. This operation takes the punch input, and computes the intersection of this input with all the images on the stack, replacing all the contents on the stack with the results. See Figure 7 for an image depicting all the changes that take place on the stack for fold operations and the punch operation.

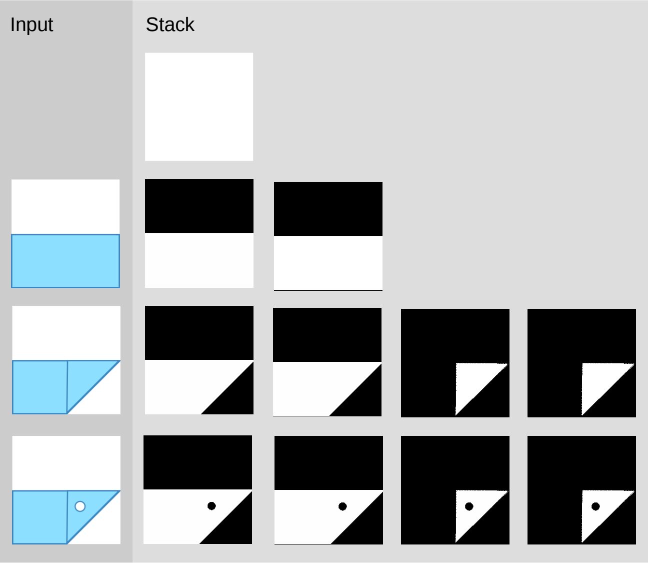

Figure 7: The various states of the image stack as inputs are fed to the model.

Unfold is the next operation after the punch has been performed. This operation runs through the operations stack and performs a reverse of all operations. It works by picking the image on top of the stack and the one below the stack. It performs the reverse operation on the image on top of the stack (which will be the folded flap).

It then performs an OR operation between the folded flap and the base image. The new image generated after the OR operation is placed onto a new image stack. The unfold operations are recursively called on the newly generated image stack until the stack contains a single image. This image will represent the model's predicted solution to the problem.

A final solution can be chosen by the model with the last image generated after the unfold operation. A pixel by pixel comparison is performed between the model's prediction and each of the possible solutions. The comparison that yields the largest number of matching pixels is selected as the solution.

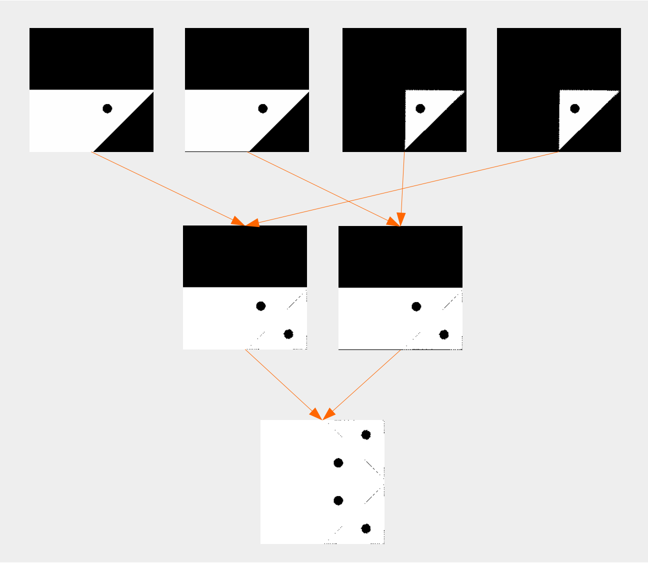

Figure 8: The various states of the image stack during the unfold operations.

Experiments and Observations

We tested the model against all twenty items from "The Punched Hole" test (Ekstrom et al., 1976). Input images (for both the fold stages and answer choices) for this experiment were taken from the original test but redrawn as "clean" versions using the Inkscape Vector Graphics editor. Redrawn vector images were converted to raster images before being passed to the model.

Results for the experiment were a count of the number of items on which the model was able to select the correct answer. When this experiment was performed, the model selected the correct answer on all items in the test—a score of 20 out of 20.

Looking more carefully at the operation of the model, we observed that a constant number of operations are performed for each fold simulated. Also, the size of the stack grows exponentially with respect to the number of folds performed. For every n folds, there are a total of 2^n items on the stack.

Interestingly, the set of problems had 1 to 3 fold levels. This meant that the maximum stack size required for the tasks varied from 2 (for single folds) to 8 (for triple folds). If we take the size of the image stack to be analogous to "working memory usage" in our model, the maximum number of items stored while solving any of the paper folding problems is consistent with what is known about visuospatial working memory capacity limits in people (Luck & Vogel, 1997).

Also, working in the image domain gives the model the ability to operate on arbitrary folding tasks. To test this ability, we ran a "paper snowflake" simulation through the model to evaluate its output. From Figure 9, we can clearly observe that the model generated an output that corresponds to the snowflake folds passed through it.

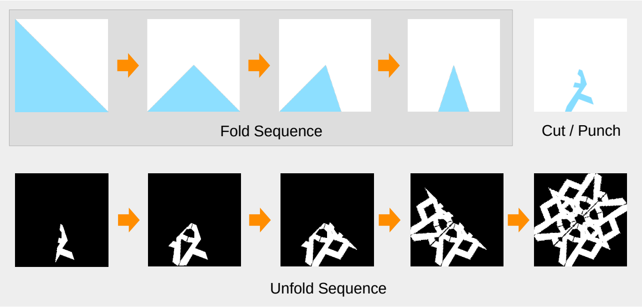

Figure 9: The model's "solution" to a paper folding problem with arbitrary fold and punch shapes. The top row of images show the input to the model and the bottom row of images showt the output of the model at the various stages of unfolding the snowflake.

Discussion and Future Work

One current assumption of this model has to do with the comparison technique used to match possible solution choices to the predicted solution. As stated earlier, this operation is performed using a pixel by pixel comparison technique, where pixels on the predicted solution must closely match pixels on a possible solution. For the case of our experiment, these possible solutions were carefully drawn such that holes in both the right and wrong choices were precisely placed.

However, as we have observed in other standardized cognitive assessments (Kunda et al., 2013; Kunda & Ting, 2015), printed figures in test booklets are not always so precise at the pixel level. (We surmise that many of these tests must have been hand-drafted when they were first created.) On the actual paper folding test by Ekstrom et al. (1976), many of the positions of the punched holes are not necessarily aligned perfectly. However, people are still able to solve these tasks, which suggests that the model should have more robust pattern recognition and processing abilities.

Our model has shown one possible set of cognitive mechanisms, based on visual mental images, that are sufficient for solving the paper folding test. Other strategies undoubtedly exist. What is more interesting, perhaps, is a consideration of why people might fail to solve paper folding items. What mechanism or set of mechanisms might they lack?

One possibility is working memory capacity, simulated in our model as the size of the image stack. Clearly, limiting the size of the stack will immediately reduce the model's ability to successfully solve paper folding problems. However, there are other possibilities as well.

One idea is that people might "forget" where the fold is on a folded up piece of mental paper, and proceed to unfold the paper in the wrong direction. (For example, they might interchange the folded side and the open side of a folded page, at the moment when they are unfolding it.) This is a subtle error that does not have to do with raw capacity but more like a limit on attentional capability, or the accurate persistence of information in working memory. We speculate that these types of fold-forgetting errors may lead participants to choose some of the *distracter* answer choices that are provided.

In continued work on the model, we will implement some of these cognitive limitations to see what might lead the model to make particular types of errors. Then, we could compare the errors made by different configurations of the model to the errors made by people, to see if there are suggestive connections between cognitive strategy variations and behavioral error patterns (Kunda, Soulières, et al., 2016).

References

Arnheim, R. (1969). *Visual thinking*. University of California Press.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2–11.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*.

Ferguson, E. S. (1994). *Engineering and the mind's eye*. The MIT press.

Foer, J. (2011). *Moonwalking with Einstein: The art and science of remembering everything*. Penguin.

Giaquinto, M. (2007). *Visual thinking in mathematics*. Oxford University Press.

Glass, L., Krueger, F., Solomon, J., Raymont, V., & Grafman, J. (2012). Mental paper folding performance following penetrating traumatic brain injury in combat veterans: a lesion mapping study. *Cerebral Cortex*, bhs153.

Keehner, M. M., Tendick, F., Meng, M. V., Anwar, H. P., Hegarty, M., Stoller, M. L., & Duh, Q.-Y. (2004). Spatial ability, experience, and skill in laparoscopic surgery. *The American Journal of Surgery*, *188*(1), 71–75.

Kunda, M. (2015). Computational mental imagery, and visual mechanisms for maintaining a goal-subgoal hierarchy. In *Proceedings of the third annual conference on advances in cognitive systems acs* (p. 4).

Kunda, M., El Banani, M., & Rehg, J. M. (2016). A computational exploration of problem-solving strategies and gaze behaviors on the block design task. In *38th annual conference of the cognitive science society, philadelphia, usa.*

Kunda, M., & Goel, A. K. (2011). Thinking in pictures as a cognitive account of autism. *Journal of autism and developmental disorders*, *41*(9), 1157–1177.

Kunda, M., McGreggor, K., & Goel, A. K. (2013). A computational model for solving problems from the Ravens Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, *22*, 47–66.

Kunda, M., Soulières, I., Rozga, A., & Goel, A. K. (2016). Error patterns on the Raven's Standard Progressive Matrices Test. *Intelligence*, *59*, 181–198.

Kunda, M., & Ting, J. (2015). Looking around the minds eye: How internal deployments of attention can affect visual search performance. In *Proc. third annual conf. advances in cognitive systems (acs).*

Lovett, A. M., & Forbus, K. D. (2013). Modeling spatial ability in mental rotation and paper-folding. In *Cogsci.*

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of educational psychology*, *95*(4), 833.

Miller, A. I. (1984). Imagery in scientific thought: Creating 20th-century physics. *Boston: Birkhauserm*.

National Research Council. (2009). Mathematics learning in early childhood: Paths toward excellence and equity. *National Academies Press*.

Petre, M., & Blackwell, A. F. (1999). Mental imagery in program design and visual programming. *International Journal of Human-Computer Studies*, *51*(1), 7–30.

Shepard, R. N., & Feng, C. (1972). A chronometric study of mental paper folding. *Cognitive psychology*, *3*(2), 228–243.

Silverman, L. K. (2002). *Upside-down brilliance: The visual-spatial learner*. DeLeon Publishing Denver, CO.

Silvia, P. J. (2008). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual differences*, *44*(4), 1012–1021.

Soulieres, I., Zeffiro, T., Girard, M., & Mottron, L. (2011). Enhanced mental image mapping in autism. *Neuropsychologia*, *49*(5), 848–857.

Thagard, P. (2005). *Mind: Introduction to cognitive science*. MIT Press.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin*, *139*(2), 352.

Verdine, B. N., Troseth, G. L., Hodapp, R. M., Dykens, E. M., & Conners, F. (2008). Strategies and correlates of jigsaw puzzle and visuospatial performance by persons with prader-willi syndrome. *American Journal on Mental Retardation*, *113*(5), 343–355.

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

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*(4), 763–771.