A Computational Exploration of Problem-Solving Strategies and Gaze Behaviors on the Block Design Task

Maithilee Kunda (mkunda@vanderbilt.edu)
Mohamed El Banani (mbanani@gatech.edu)
James M. Rehg (rehg@gatech.edu)
School of Interactive Computing, Georgia Institute of Technology
85 Fifth Street NW, Atlanta, GA 30332 USA

Abstract

The block design task, a standardized test of nonverbal reasoning, is often used to characterize atypical patterns of cognition in individuals with developmental or neurological conditions. Many studies suggest that, in addition to looking at quantitative differences in block design speed or accuracy, observing qualitative differences in individuals' problem-solving strategies can provide valuable information about a person's cognition. However, it can be difficult to tie theories at the level of problem-solving strategy to predictions at the level of externally observable behaviors such as gaze shifts and patterns of errors. We present a computational architecture that is used to compare different models of problem-solving on the block design task and to generate detailed behavioral predictions for each different strategy. We describe the results of three different modeling experiments and discuss how these results provide greater insight into the analysis of gaze behavior and error patterns on the block design task.

Keywords: artificial intelligence; cognitive assessment; nonverbal intelligence; spatial reasoning; visual attention.

Introduction

The block design task is a cognitive assessment that is commonly used to measure nonverbal intelligence. In this task, a person has to reconstruct a given printed design using red and white blocks, as shown in Fig. 1. Originally devised in 1920 (Kohs, 1920), block design has been included in a multitude of standard neuropsychological test batteries, including every single edition of the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Adult Intelligence Scale (WAIS), which are among the most commonly used cognitive assessments around the world (Wechsler, 2003).

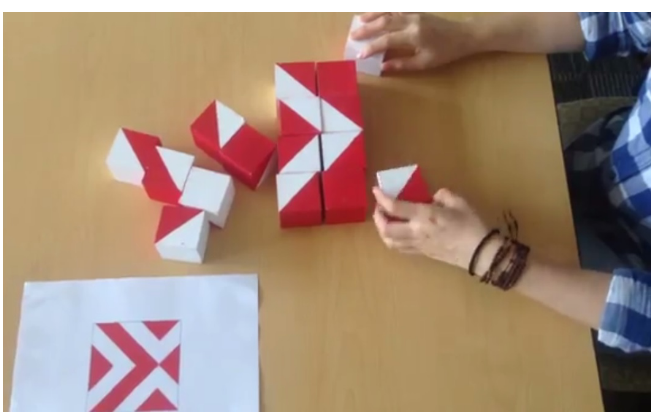

Figure 1: A person solving an example block design item. (To protect test security, actual test items are not shown.)

In the typical population, block design performance shows quantitative improvements with age in children (Wechsler, 2003) and a declining trend in the elderly (Ardila & Rosselli, 1989; Rönnlund & Nilsson, 2006). In addition to its value in the detection of brain damage (Lezak, 2004), block design is also sensitive to atypical development. Block design represents an area of strength for many individuals on the autism spectrum (Caron, Mottron, Berthiaume, & Dawson, 2006; Shah & Frith, 1993) and is an area of weakness in many individuals diagnosed with Williams syndrome (Farran, Jarrold, & Gathercole, 2001; Hoffman, Landau, & Pagani, 2003).

Kohs devised block design as a language-independent alternative to the Binet intelligence scale, which loaded heavily on verbal abilities (Kohs, 1920). The original scoring system takes into account the successful reproduction of the design as well as the time taken by the participant and the number of moves. Number of moves was dropped in later standardizations as being too cumbersome for daily practice (Hutt, 1932); current scoring schemes use final success and reaction time as the primary outcome measures. However, many researchers observe that obtaining more detailed measures of block design performance can add value by pinpointing specific cognitive processes and problem-solving strategies that an individual is employing to solve the task.

In particular, many studies have proposed that qualitative differences in block design performance are as important as quantitative differences in accuracy or reaction time. Several studies have observed that patterns of visual attention, i.e. eye gaze, between the original design and the design under construction may serve as markers for certain cognitive processes that are meaningful to the final outcome (Hoffman et al., 2003; Rozencwajg & Corroyer, 2002; Rozencwajg et al., 2005; Rozencwajg & Fenouillet, 2012).

Observing the errors that people make can also provide valuable information. Errors have been studied in terms of the particular sequence of moves that a participant makes (Joy, Fein, Kaplan, & Freedman, 2001; Toraldo & Shallice, 2004), any incorrect placements of blocks (Ben-Yishay, Diller, Mandleberg, Gordon, & Gerstman, 1971; Hoffman et al., 2003; Jones & Torgesen, 1981; Joy et al., 2001; Schatz, Ballantyne, & Trauner, 2000; Troyer, Cullum, Smernoff, & Kozora, 1994), the qualitative type or scale of errors that are made (Akshoomoff, Delis, & Kiefner, 1989; Joy et al., 2001), and also what is termed a "broken configuration" error, in which the participant places a block too far away from their current construction, violating implied configural outlines of the given design (Akshoomoff et al., 1989; Akshoomoff & Stiles, 1996; Joy et al., 2001; J. Kramer, Blusewicz, Kaplan, & Preston, 1991; J. H. Kramer, Kaplan, Share, & Huckeba, 1999; Schatz et al., 2000; Troyer et al., 1994; Zipf-Williams, Shear, Strongin, Winegarden, & Morrell, 2000).

However, missing from these accounts is a precise description of the information processing mechanisms involved in solving block design items. As a result, we also know relatively little about how these mechanisms emerge in typical development or how they are altered in atypical development.

Obtaining a better understanding of the specific cognitive processes underlying performance on standard neuropsychological assessments, especially taking into account both quantitative and qualitative differences, provides value in understanding the basic science of cognition and development as well as the interpretation of these assessments in practical settings (Hunt, 1983; Kaplan, 1988; Keating & Bobbitt, 1978; Mislevy & Verhelst, 1990; Sternberg, 1988). Computational modeling is an excellent method for this kind of detailed scientific inquiry. Implementing a computational model forces precision in the theory being tested. Royer has argued for increased specificity in theories of block design performance through conceptual information processing models (Royer, 1984), and having a working, runnable computational model takes this desired level of specificity one step further.

We propose a new computational architecture for modeling the process of solving a block design item as a recurring interplay between attention, perception, memory, and action. We implemented this architecture using a custom block design simulation environment. We present the results from three computational experiments using this architecture and discuss how the architecture can be used to better understand the relationships between different problem-solving strategies and the gaze behaviors and errors that they generate.

There are several existing cognitive architectures used to model human visuospatial task performance, such as ACT-R (Gunzelmann & Lyon, 2011) and SOAR (Laird, 2008). However, these architectures were originally developed using frameworks of cognition that prioritize symbolic processing. Modules for representing and manipulating nonsymbolic information, for instance mental imagery, have been incorporated into these architectures relatively recently. In our own work, we study mechanisms of intelligent problem solving in which visual mental images form the primary (and often only) mode of representation, and we have developed a series of computational models that adopt this perspective to solve different neuropsychological assessments, such as the Raven's Progressive Matrices test (Kunda, McGreggor, & Goel, 2013) and the Embedded Figures test (Kunda & Ting, 2015).

We have built upon this prior work in the design of the block design architecture presented in this article. The main reason that we chose to construct a new architecture instead of using an existing framework like ACT-R or SOAR is that, because our research prioritizes the role of perceptual mental representations in problem solving, we have designed our architecture to likewise rely primarily on perceptual representations, instead of having a symbolic representation system that is common to a variety of perceptual representation modules.

The Simulation Environment

We have developed a computational architecture that can autonomously solve the block design task in a simulated environment. The current environment is simplified in many ways; ongoing refinements to the environment and architecture will gradually introduce more realistic representations.

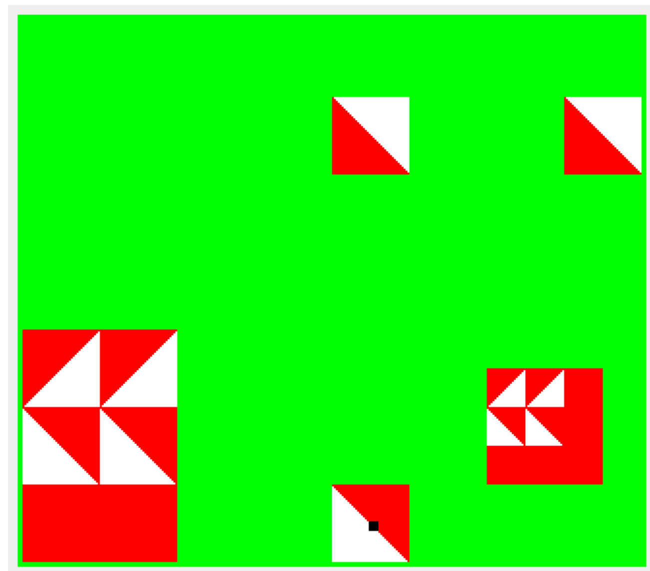

Figure 2: The simulated environment used by our architecture, representing a top-down view of the tabletop containing all of the materials for one block design item. The black dot in the image (bottom center) indicates the simulated gaze location of the architecture at the current time step.

The perceptual input available to the model consists of an overhead image of the table, as shown in Fig. 2. The table image is divided into the following regions:

1. **The target design**: The target design remains fixed for the duration of each test item. The design is represented by a single image scaled down to half the size of the actual constructed designs, consistent with standard block design administration procedures (Kohs, 1920; Wechsler, 2003).
2. **The construction area**: The construction area is located at the bottom left corner of the table. This area is empty at time 0, and has the ability to contain an assortment of blocks placed at various locations. This part of the environment is simplified by having an implied rectilinear grid that only allows blocks to be placed in slots that comply with this grid. As a result, block misalignment errors while performing the task cannot be replicated within the current generation of models. However, broken configuration errors can be replicated, as the construction area does not

have any explicit borders, allowing the model to assemble a design that extends beyond the implied boundaries.

3. **The block bank**: The block bank is the area located at the top right corner of the table and is similar to the construction area in that it contains an assortment of blocks placed at various locations and in various orientations. This area contains all of the available blocks at time 0.
4. **The hand area**: This area of the table represents the block being manipulated by the model and is located to the right of to the construction area on the table. At each time step, this area can either contain a block or be empty.
5. **The set of blocks**: The table contains N blocks, with at least as many blocks as are required to construct the target design. Each block has a 2-D position as well as a 3-D orientation to indicate the top-most face of the block.

The behavioral variables that can be captured from this environment include: success/failure, sequence of block moves, and gaze shifts to each area of the environment. While the simulated environment is capable of supporting time-based measurements of model performance, detailing timing information is being incorporated into the architecture as part of ongoing work and is not used in the current study.

The Computational Architecture

We designed the block design computational architecture to have six basic mechanisms, as shown in Fig. 3:

1. **The visual perception module** takes as input a visual scene representation from the simulation and outputs a new image at each time step, which we call the perceptual image. The content of this image varies depending on where the model is currently directing its gaze. The perceptual image has a fixed resolution but can "zoom" in or out relative to the table, which allows the model to capture different amounts of visual information from the scene, ranging from the general form of a large section of the table to a highly detailed image of a small section of the table.
2. **The visual attention module** specifies a gaze location at each time step, which can be directed to specific points within one of four broad locations: the target design, the available blocks area, the design being constructed, and the hand location. This module also includes mechanisms for performing visual search either for any block or for a block whose top-most face best matches the block face stored in the short-term memory buffer.
3. **The mental imagery module** stores the perceptual image in a short-term memory buffer for later use as a "mental image." The mental imagery module can compute a measure of visual similarity between the current perceptual image and the current mental image (Kunda et al., 2013).
4. **The short term memory buffer** stores the output of the mental imagery module as well as any relevant spatial information that is required to solve the task. In the current implementation, the only spatial information that is stored is a reference point on the current object/area of interest, in the form of Cartesian coordinates.
5. **The motor action module** takes actions that change the state of the environment. The available actions include: 1) selecting a particular block from the table and putting this block into the hand location, 2) rotating a block that is in the hand location, the block bank, or the construction area, or 3) placing a block currently held in the hand to a specific position inside the construction area or the block bank.
6. **The central executive module** is responsible for planning and decision making, and it coordinates the overall operation of the model. At each time step, this module can execute any action from a set of available actions. The available actions include changing the gaze location, performing a visual search for a specific image, storing, manipulating, or flushing the mental image, and selecting, rotating, or placing a block in the simulated environment.

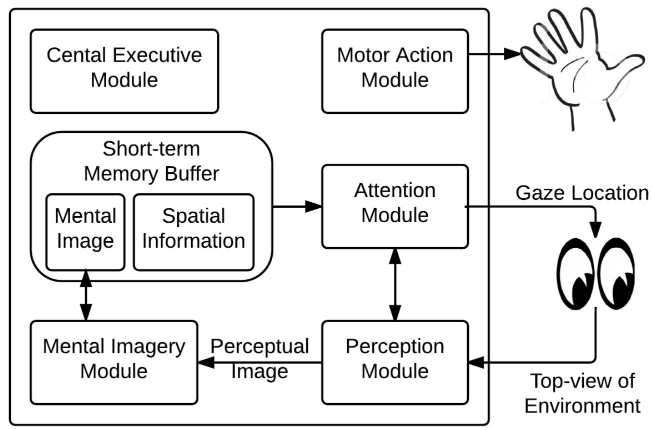

Figure 3: The overall model architecture, including inputs and outputs to the simulated block design task environment described in the previous section. The central executive module has bidirectional connections to the other five modules in the system; for clarity, these connections are not shown.

Experiments and Results

Using the simulation environment described above, we ran three computational experiments in which we compared the performance of different block design models built using the same underlying architecture. Each model was presented with 9 blocks and a single 3x3 target design, which appears in the lower right corner of Fig. 2. We present results from the model averaged over 1000 runs. Randomness effects within a single run stem from the initial random positioning of the blocks as well as randomness in the model's search strategies.

While there are likely many high-level strategies that can be used to solve block design items, the current models that we tested all follow what has been termed the "analytic strategy," in which "the displayed design is mentally segmented into units corresponding to block faces, then the blocks are directly placed, one by one, to match each unit" (Schorr, Bower, & Kiernan, 1982). Other high-level strategies, including the influence of Gestalt effects on perception, are not explicitly implemented in the current family of models but are part of ongoing work to improve and expand the basic architecture.

Because there are random factors in the environment and in each model's processing, we measured the output of the models by constructing gaze transition graphs. These graphs show the average number of gaze transitions made by each model within or between four different areas of the simulation environment: the block bank, the construction area, the target design, and the model's "hand."

Experiment #1: Visual Working Memory

In solving a block design item, each model first looks at the design, stores some part of the design in memory, then selects and uses blocks to reconstruct this stored part of the design. Then, the model returns to the design to see and store the next part of the design. This experiment investigates the effects of the size of a model's visual working memory on its task performance, in terms of how much information from the design can be stored at each iteration.

Each model stores some part of the design in its mental imagery module. In the current implementation, the mental images are square and have a fixed size, defined by the number of blocks they can store. We created two different models, one with a 1x1 mental image size and one with a 3x3 mental image size. The mental images are stored at the same resolution at which they are perceived.

Each of the two models applied a guided search strategy to locate the blocks and used a visual similarity threshold of 0.95. (These two additional parameters are discussed in detail in the next two experiments.) The gaze transition graphs for each of these models are shown in Fig. 4.

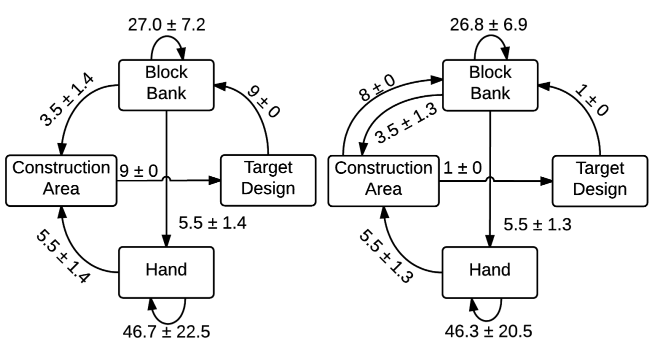

Figure 4: Gaze transitions made by models with a mental image size of 1x1 (left) or 3x3 (right). Results show *mean +/- standard deviation* over 1000 trials.

Two noticeable effects can be seen in this figure. First, the model with larger mental image capacity only has to look at the target design once, while the other model must return its gaze to the target design many times. Second, the model with larger mental image capacity shows more gaze transitions between the block bank and construction area, because it is working off a mental image instead of having to return to the design after placing each block.

Experiment #2: Search Strategy

Each model has a search strategy that determines which block from the block bank is chosen to add to the construction area. Because the blocks are identical, a random choice can always suffice. In the *random search strategy*, the model first randomly chooses a block from the block bank and moves it to the hand area. Then, the model segments the target design to determine the block face that should be added next to the construction area. If the top-most face of the block in hand matches the target block face, the block is immediately added to the construction area. Otherwise, the block is rotated to match the target before being added to the construction area.

Alternately, a model can be more strategic about its choice from the block bank. In the *guided search strategy*, the model first segments the target design to determine the block face that should be added next to the construction area. The model then proceeds to search the remaining available blocks in the block bank for the target block face. If an exact match is found, the matching block is added to the construction area. If not, the best match is picked up and rotated to match the target block face and then added to the construction area.

Both models have a mental image size of 1x1 and a visual similarity threshold of 0.95. The gaze transition graphs for each of these models are shown in Fig. 5.

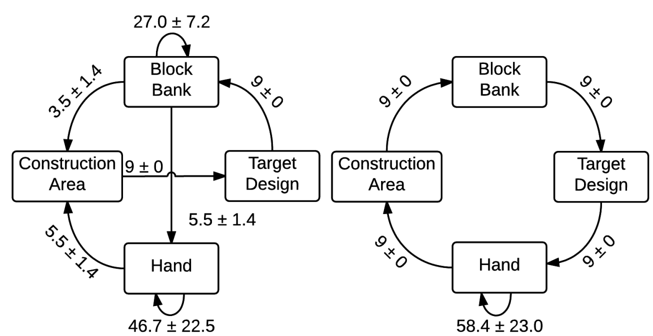

Figure 5: Gaze transitions made by models using a guided search strategy (left) versus a random search strategy (right). Results show *mean +/- standard deviation* over 1000 trials.

In this experiment, it is interesting to see that the random search strategy is, in a sense, more efficient. Gaze transitions follow a very definite pattern, and the only added gaze behaviors are those that monitor the rotation of each block while it is being held. The model following the guided search strategy spends much more time looking through the blocks in the block bank, and even then, the model still must monitor a large number of in-hand block rotations.

Experiment #3: Visual Similarity Threshold

The models use a measure of visual similarity to compare two images, for example to compare a mental image to a perceived block in the environment. Each model has a similarity threshold to determine when two images are considered to be the same. In the current implementation, visual similarity is calculated simply as the proportion of matching pixels between two images (where 0.0 indicates no match and 1.0 indicates a perfect match). We varied this visual similarity threshold to be 0.95 in one model and 0.64 in another.

Both models have a mental image size of 1x1 and use a guided search strategy. The gaze transition graphs for each of these models are shown in Fig. 6.

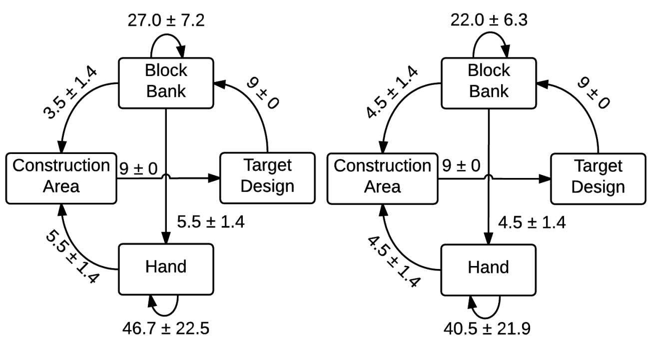

Figure 6: Gaze transitions made by models having a visual similarity threshold of 0.95 (left) or 0.64 (right). Results show *mean +/- standard deviation* over 1000 trials.

Both of these graphs show the same connectivity. However, the model with the lower similarity threshold value shows fewer gaze behaviors on both the block bank (while searching for a matching block) and on the hand (while rotating blocks as needed). This decrease in visual effort makes sense given that the model is more lenient with what it considers to be a visual match.

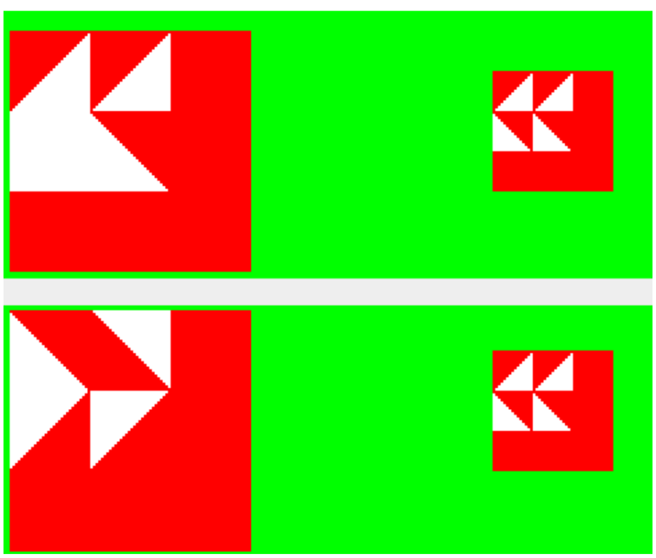

Figure 7: Examples of errors made by models with visual similarity thresholds of 0.64 (top) and 0.5 (bottom).

Lowering the similarity threshold can also lead to errors. While the model with the higher threshold makes no errors, the model with the lower threshold produces many incorrect reconstructions of the target design, with overall accuracy falling to 25.4%. The errors produced by decreasing the visual similarity threshold are specific to the incorrect placement of blocks that share high visual similarity with the correct block that should have been placed at that position. Two examples of the kinds of errors that occur at different threshold values are shown in Fig. 7. At a threshold value of 0.64, the errors take the form of placing a solid face in the place of a diagonal one. At a lower threshold value of 0.5, a model could replace a diagonal face block with a diagonal face block that has a different 2-D orientation. However, more acute errors, such as a solid red face being replaced by a solid white face or a diagonal face being replaced by its complement do not occur even with a similarity threshold of 0.5, as the visual similarity between such extreme pairs is much lower.

Discussion and Future Work

Despite the ubiquity of block design in clinical, scientific, and educational settings, we still do not know the precise nature of the mechanisms and strategies that people use to solve the task. Describing these mechanisms and strategies at a computational level provides an experimental platform on which to construct and flesh out theories of problem-solving on the block design task and to make detailed predictions about behavior, including both quantitative and qualitative variations.

We have shown how computational models of problem-solving on the block design task can be used to test the effects that different variations in strategy and other cognitive characteristics have on behavior, including accuracy, gaze patterns, and the types of errors that are made. While the models we have presented are admittedly simplified in comparison to human performance, these models give us a structured way to understand both the requirements of the task and how specific mechanisms relate to specific behaviors.

In future work, we will collect data from human participants to analyze their gaze transition graphs and error patterns, comparing them to those of our computational models, to gain more insight into the variability of strategy and behavior across our participant sample. These analyses will provide insight into cognitive development and individual differences in typically developing children as well as in individuals with different developmental or neuropsychological conditions.

Continued work in this direction will: 1) enhance our scientific understanding of how nonverbal intelligence develops and matures, 2) enable the generation of detailed behavioral predictions and research questions for future scientific inquiry, 3) inform the practical development and interpretation of neuropsychological assessments, and 4) provide a conceptual bridge to map behavioral observations onto measurements and models of neural activity.

Acknowledgments

We thank Y. Kang, M. Munch, E. Nkadi, and R. Stauffer for contributions to this project. This research was supported by NSF Expedition Award #1029679 and by the Georgia Tech Undergraduate Research Opportunities Program.

References

Akshoomoff, N. A., Delis, D. C., & Kiefner, M. G. (1989). Block constructions of chronic alcoholic and unilateral brain-damaged patients: A test of the right hemisphere vulnerability hypothesis of alcoholism. *Archives of Clinical Neuropsychology*, *4*(3), 275–281.

Akshoomoff, N. A., & Stiles, J. (1996). The influence of pattern type on children's block design performance. *J. International Neuropsychological Society*, *2*(5), 392–402.

Ardila, A., & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. *Dev. Neuropsych.*, *5*, 307–320.

Ben-Yishay, Y., Diller, L., Mandleberg, I., Gordon, W., & Gerstman, L. (1971). Similarities and differences in block design performance between older normal and brain-injured persons: A task analysis. *J. Abnormal Psychology*, *78*, 17-25.

Caron, M., Mottron, L., Berthiaume, C., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinnings of visuospatial peaks in autism. *Brain*, *129*(7), 1789–1802.

Farran, E. K., Jarrold, C., & Gathercole, S. E. (2001). Block design performance in the Williams syndrome phenotype: A problem with mental imagery? *J. Child Psychology and Psychiatry*, *42*(6), 719–728.

Gunzelmann, G., & Lyon, D. R. (2011). Representations and processes of human spatial competence. *Topics in Cognitive Science*, *3*(4), 741–759.

Hoffman, J. E., Landau, B., & Pagani, B. (2003). Spatial breakdown in spatial construction: Evidence from eye fixations in children with Williams syndrome. *Cognitive Psychology*, *46*(3), 260–301.

Hunt, E. (1983). On the nature of intelligence. *Science*, *219*(4581), 141–146.

Hutt, M. (1932). The Kohs block-design tests: A revision for clinical practice. *J. Applied Psychology*, *16*(3), 298-307.

Jones, R. S., & Torgesen, J. K. (1981). Analysis of behaviors involved in performance of the block design subtest of the WISC-R. *Intelligence*, *5*(4), 321–328.

Joy, S., Fein, D., Kaplan, E., & Freedman, M. (2001). Quantifying qualitative features of block design performance among healthy older adults. *Archives of Clinical Neuropsychology*, *16*(2), 157–170.

Kaplan, E. (1988). The process approach to neuropsychological assessment. *Aphasiology*, *2*(3-4), 309–311.

Keating, D. P., & Bobbitt, B. L. (1978). Individual and developmental differences in cognitive-processing components of mental ability. *Child Development*, *49*(1), 155–167.

Kohs, S. (1920). The block-design tests. *J. Experimental Psychology*, *3*(5), 357-376.

Kramer, J., Blusewicz, M., Kaplan, E., & Preston, K. (1991). Visual hierarchical analysis of block design configural errors. *J. Clin. and Exp. Neuropsychology*, *13*(4), 455–465.

Kramer, J. H., Kaplan, E., Share, L., & Huckeba, W. (1999). Configural errors on WISC-III block design. *J. International Neuropsychological Society*, *5*(6), 518–524.

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., & 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).*

Laird, J. E. (2008). Extending the Soar cognitive architecture. In *Proc. 2008 Conf. Artificial General Intelligence (AGI).*

Lezak, M. D. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*(2), 195–215.

Rönnlund, M., & Nilsson, L.-G. (2006). Adult life-span patterns in WAIS-R block design performance: Cross-sectional versus longitudinal age gradients and relations to demographic factors. *Intelligence*, *34*(1), 63–78.

Royer, F. L. (1984). Stimulus variables in the block design task: A commentary on Schorr, Bower, and Kiernan. *J. Consulting and Clinical Psychology*, *52*(4), 700-704.

Rozencwajg, P., Cherfi, M., Ferrandez, A., Lautrey, J., Lemoine, C., & Loarer, E. (2005). Age related differences in the strategies used by middle aged adults to solve a block design task. *Int. J. Aging & Human Dev.*, *60*(2), 159–182.

Rozencwajg, P., & Corroyer, D. (2002). Strategy development in a block design task. *Intelligence*, *30*(1), 1–25.

Rozencwajg, P., & Fenouillet, F. (2012). Effect of goal setting on the strategies used to solve a block design task. *Learning and Individual Differences*, *22*(4), 530–536.

Schatz, A., Ballantyne, A., & Trauner, D. (2000). A hierarchical analysis of block design errors in children with early focal brain damage. *Dev. Neuropsychology*, *17*(1), 75–83.

Schorr, D., Bower, G. H., & Kiernan, R. J. (1982). Stimulus variables in the block design task. *J. Consulting and Clinical Psychology*, *50*(4), 479-487.

Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *J. Child Psychology and Psychiatry*, *34*(8), 1351–1364.

Sternberg, R. J. (1988). Applying cognitive theory to the testing and teaching of intelligence. *Applied Cognitive Psychology*, *2*(4), 231–255.

Toraldo, A., & Shallice, T. (2004). Error analysis at the level of single moves in block design. *Cognitive Neuropsychology*, *21*(6), 645–659.

Troyer, A., Cullum, C., Smernoff, E., & Kozora, E. (1994). Age effects on block design: Qualitative performance features and extended-time effects. *Neuropsychol.*, *8*, 95-99.

Wechsler, D. (2003). *Wechsler intelligence scale for children (WISC-IV)*. The Psychological Corporation.

Zipf-Williams, E., Shear, P., Strongin, D., Winegarden, B., & Morrell, M. (2000). Qualitative block design performance in epilepsy patients. *Archives of Clinical Neuropsychology*, *15*, 149–157.