

Taking a Look (Literally!) at the Raven's Intelligence Test: Two Visual Solution Strategies

Maithilee Kunda (mkunda@gatech.edu)
Keith McGregor (keith.mcgreggor@gatech.edu)
Ashok Goel (goel@cc.gatech.edu)

Design & Intelligence Laboratory, School of Interactive Computing, Georgia Institute of Technology
85 Fifth Street NW, Atlanta, GA 30332 USA

Abstract

The Raven's Progressive Matrices intelligence test is widely used as a measure of Spearman's general intelligence factor g . Although Raven's problems resemble geometric analogies, prior computational accounts of solving the test have been propositional. Studies of both typical and atypical human behavior suggest the possible existence of visual strategies; for example, neuroimaging data indicates that individuals with autism may preferentially recruit visual processing brain regions when solving the test. We present two different algorithms that use visual representations to solve Raven's problems. These algorithms yield performances on the Standard Progressive Matrices test at levels equivalent to typically developing 9.5- and 10.5- year-olds. We find that these algorithms perform most strongly on problems identified from factor-analytic human studies as requiring gestalt or visuospatial operations, and less so on problems requiring verbal reasoning. We discuss implications of this work for understanding the computational nature of Raven's and visual analogy in problem solving.

Keywords: Analogy; intelligence tests; knowledge representations; mental imagery; Raven's Progressive Matrices; visual reasoning.

Introduction

The Raven's Progressive Matrices tests (Raven, Raven, & Court, 1998) are a collection of standardized intelligence tests that consist of geometric analogy problems in which a matrix of geometric figures is presented with one entry missing, and the correct missing entry must be selected from a set of answer choices. Figure 1 shows an example of a 2x2 matrix problem that is similar to one in the Standard Progressive Matrices (SPM); other problems contain 3x3 matrices. The entire SPM consists of 60 problems divided into five sets of 12 problems each (sets A, B, C, D & E), roughly increasing in difficulty both within and across sets.

Although the Raven's tests are supposed to measure only educative ability, or the ability to extract and understand information from a complex situation (Raven, Raven, & Court 1998), their high level of correlation with other multi-domain intelligence tests have given them a position of centrality in the space of psychometric measures (e.g. Snow, Kyllonen, & Marshalek 1984), and as a result, they are often used as tests of general intelligence in clinical, educational, vocational, and scientific settings.

Computational accounts of problem solving on the Raven's tests have, with the exception of Hunt (1974), assumed that visual inputs are translated into propositions,

over which various kinds of reasoning then take place. In this paper, we provide evidence from two different methods that Raven's problems can be solved visually, without first converting problem inputs into propositional descriptions.

Existing Computational Accounts

Carpenter, Just, and Shell (1990) used a production system that took hand-coded symbolic descriptions of problems from the Advanced Progressive Matrices (APM) test and then selected an appropriate rule to solve each problem. The rules were generated by the authors from *a priori* inspection of the APM and were validated in experimental studies of subjects taking the test with verbal reporting protocols. Bringsjord and Schimanski (2003) used a theorem-prover to solve selected Raven's problems stated in first-order logic.

Lovett, Forbus, and Usher (2007) combined automated sketch understanding with the structure-mapping analogy technique to solve problems from the Standard Progressive Matrices (SPM) test. Their system took as inputs problem entries sketched in Powerpoint as segmented shape objects and then automatically translated these shapes into propositional descriptions. A two-stage structure-mapping process was then used to select the answer that most closely fulfilled inferred analogical relations from the matrix.

In contrast to these propositional approaches, Hunt (1974) proposed the existence of two qualitatively different strategies: "Gestalt," which used visual representations and perceptual operations like continuation and superposition,

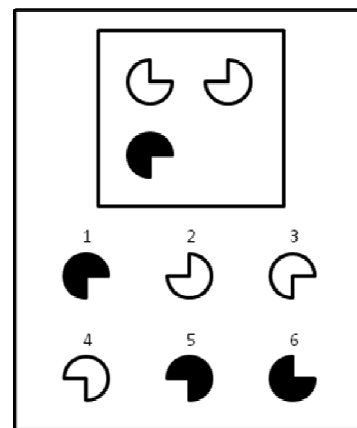


Figure 1: Example problem similar to one from the Standard Progressive Matrices (SPM) test.

and “Analytic,” which used propositional representations and logical operations. The Analytic algorithm is similar to that of Carpenter, Just, and Shell (1990) in that it applied rules to lists of features representing each matrix entry. The Gestalt algorithm is similar to our methods in that it used visual operations over imagistic problem inputs, but it differs in that it operated on the entire problem matrix as a single image, whereas our methods treat each matrix entry as a separate image. While Hunt’s algorithms provide an intuitively appealing account of solving Raven’s problems, neither algorithm was actually implemented.

Behavioral Evidence for Multiple Strategies

Studies of human behavior suggest that qualitatively distinct problem solving strategies can be used to solve Raven’s problems. Factor analyses of both the SPM (Lynn, Allik, & Irving, 2004; van der Ven & Ellis, 2000) and the APM (Dillon, Pohlmann, & Lohman, 1981; Mackintosh & Bennett, 2005; Vigneau & Bors, 2005) have identified multiple factors underlying these tests, which often divide test problems into two categories: those solvable using visuospatial or gestalt operations and those solvable using verbal reasoning. In support of this dichotomy, DeShon, Chan, and Weissbein (1995) found that simultaneously performing a verbal overshadowing protocol differentially impaired accuracy on about half of APM problems.

These studies of typically developing individuals have generally focused on within-individuals differences in solution strategies, i.e. a particular individual using different strategies on different portions of the test in a single sitting. Recent evidence from autism offers evidence of between-individuals strategy differences as well: individuals with autism do not show the same correlations between Raven’s scores and other cognitive measures that are robustly demonstrated by typically developing individuals (Dawson, Soulières, Gernsbacher, & Motttron, 2007).

Even more striking are recent neuroimaging data that show increased brain activation in visual regions for individuals with autism solving the SPM than controls (Soulières et al., 2009). This study also found significant differences in reaction time as a function of problem type, with problems classified as “figural” or “analytic” based on previously published factor-analytic studies. The results from this study are highly suggestive of individuals with autism using a visual strategy that contrasts with the strategy used by controls. Evidence for a visual strategy preference in autism is found across several other cognitive task domains as well (Kunda & Goel, 2008).

Our approach

We hypothesize that Raven’s problems can be solved computationally using purely visual representations. To test this hypothesis, we have developed two different algorithms that in this paper we will call the “affine” method and the “fractal” method. Both methods use image transformations to solve Raven’s problems *without* converting the input images into any kinds of propositions. Below, we describe

each of these algorithms, followed by an analysis of their performance on all 60 problems from the Raven’s Standard Progressive Matrices (SPM) test.

Visual Methods for the Raven’s Test

Similitude Transformations

At the core of each of our algorithms are image operations that fall under the category of affine transformations, and in particular similarity-preserving or “similitude” transforms. Similitude transforms can be represented as compositions of dilation (i.e. scaling), orthonormal transformation, and translation. Our implementations presently examine the identity transform, horizontal and vertical reflections, and 90°, 180°, and 270° orthonormal rotations, composed with various translations. The affine method restricts dilation to a value of one, i.e. no scaling, whereas the fractal method uses a short sequence of progressively smaller dilation values, i.e. its similitude transformations are contractive.

There is evidence that human visual processing can apply some of these types of transformations to mental images, or at least operations that are computationally isomorphic in some sense. In the theory of mental imagery proposed by Kosslyn, Thompson, and Ganis (2006), transformations of mental images include scanning (i.e. translation), zooming (i.e. scaling), and rotation, among others.

A Model of Similarity

Similarity lies at the core of both of our accounts of visual problem solving on the Raven’s test. We calculate visual similarity using the ratio model (Tversky, 1977):

$$similarity(A, B) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \quad (1)$$

In this equation, f represents some function over features in each of the specified sets; for instance, f might simply be a count of features. The constants α and β are used as weights for the non-intersecting portions of the sets A and B . If α and β are both set to one, then this equation becomes:

$$similarity(A, B) = \frac{f(A \cap B)}{f(A \cup B)} \quad (2)$$

Equation (2) is used in both the affine and fractal methods, and it yields maximal similarity for sets in which A is equal to B . In contrast, if α is set to one and β is set to zero, it yields maximal similarity for sets in which A is a proper subset of B . If α is set to zero and β is set to one, then the opposite holds, and maximal similarity is found for sets in which B is a proper subset of A . These two variants are used in the affine method to capture notions of image composition, i.e. image addition and subtraction.

In the affine method, each feature is defined as a pixel, and intersection, union, and subtraction operations are defined as the product, maximum, and difference of RGB pixel values, respectively. The fractal method uses features derived from different combinations of elements from the fractal encoding (McGregor, Kunda, & Goel, 2010).

The Affine Method

The affine method assumes that elements within a row or column in a Raven's problem matrix are related by similitude transformations. It tries to discover which similitude transformation best fits any of the complete rows or columns in the matrix, and then applies this transform to the last row/column to generate a guess for the answer. Then, it compares this guess to each of the answer choices, and chooses the answer that is most similar.

Each similitude transformation is represented as the combination of three image operations: a base transform, a translation, and a composition. Algorithm 1 shows how, for a pair of images A and B, these three components of the "best-fit" similitude transformation are found. Given a Raven's problem, then, the affine method seeks to discover the best-fit similitude transform over various combinations of the matrix entries. In particular, the algorithm assumes that certain analogical relationships exist based on the spatial arrangement of the entries. Similitude transforms are calculated for those combinations of entries that would yield an analogical mapping to solve for the missing entry. The specific base transforms and analogical relationships used by the affine algorithm are shown in Table 1, divided into those used for 2x2 and for 3x3 matrix problems.

Once the relationship and transformation are found that maximize similarity, the transformation is applied to the first entry or entries in the last row or column, as listed in Table 1. The resulting image represents the algorithm's best guess as to the missing entry. This image is compared to the answer choices, using Equation (2), and the best match is chosen as the final answer.

For each base transform T:

Apply T to Image A.

Find translation (tx, ty) which yields best match between T(A) and B, using Eq. (2).

Find image composition operand X as follows:

Calculate similarity using Eq. (1) with:

1) $\alpha = 1, \beta = 1$

2) $\alpha = 1, \beta = 0$

3) $\alpha = 0, \beta = 1$

Choose maximum similarity value.

If maximum is (1), then $X = 0$.

If maximum is (2), then $X = B - A$,

and \oplus refers to image addition.

If maximum is (3), then $X = A - B$,

and \oplus refers to image subtraction.

The best-fit similitude transformation can then be specified as:

$$[T_{\max} + (tx, ty)](A) \oplus X = B$$

Algorithm 1. Affine method for calculating best-fit similitude transformation for a pair of images A and B. For three-element transforms, T is applied to images A and B, and the result is compared, as above, to image C.

Table 1: Base transforms and matrix relationships used by the affine algorithm.

Transforms	2x2:		3x3:		
	A	B	A	B	C
	C	?	D	E	F
			G	H	?
Two-element transforms & relations	Identity		BC→H?	DG→F?	
	Mirror		AC→G?	AG→C?	
	Flip	AB→C?	EF→H?	EH→F?	
	Rotate90	AC→B?	DF→G?	BH→C?	
	Rotate180		GH→H?	CF→F?	
	Rotate270				
Three-element transforms & relations	Union		ABC→GH?		
	Intersection	n/a	DEF→GH?		
	XOR		ADG→CF?		
			BEH→CF?		

For example, take the problem given in Figure 1. The similarity scores calculated for the various transforms and relationships are shown in Table 2. The best-fit similitude transformation is found to be a mirror (or reflection about the vertical axis) for the relationship AB, using an addition image composition (i.e. maximal similarity found using $\alpha = 1, \beta = 0$). Therefore, the answer image "?" is obtained using the analogous relationship of A:B :: C?: C is mirrored, translated by the (tx, ty) that was found in the search, and the composition operand of B - A (which in this case is mostly a blank image) is added on to the result. Finally, this "guess" image is compared to each of the six answer choices using Equation (2), and the best match is chosen as the final answer, which in this case is answer #5.

Table 2: Calculation of best-fit similitude transform and resulting answer guess for the problem shown in Figure 1.

Relation	Transform	$\alpha = 1$ $\beta = 1$	$\alpha = 1$ $\beta = 0$	$\alpha = 0$ $\beta = 1$
AB	Identity	0.475	0.644	0.644
	Mirror	0.963	0.981	0.981
	Flip	0.337	0.504	0.504
	Rotate90	0.341	0.508	0.508
	Rotate180	0.453	0.624	0.624
	Rotate270	0.947	0.973	0.973
AC	Identity	0.256	0.764	0.277
	Mirror	0.252	0.759	0.274
	Flip	0.335	0.951	0.341
	Rotate90	0.331	0.941	0.338
	Rotate180	0.257	0.771	0.279
	Rotate270	0.250	0.752	0.273

Generated guess:



The Fractal Method

The fractal method proceeds in a manner which at once resembles and yet differs from the affine method. Like the affine method, the fractal method seeks to find a re-representation of the images within a Raven's problem as a set of similitude transformations. Unlike the affine method, the fractal method seeks these representations at a significantly finer partitioning of the images, and uses these representations (and more precisely, features derived from these representations) to determine similarity for each possible answer, simultaneously, across the bulk of relationships present in the problem.

The mathematical derivation for the process of fractal image representation expressly depends upon the notion of real world images, i.e. images that are two dimensional and continuous (Barnsley & Hurd, 1992). Two key observations are that all naturally occurring images we perceive appear to have similar, repeating patterns, and, no matter how closely we examine the real world, we find instances of similar structures and repeating patterns. These observations suggest that it is possible to describe the real world in terms other than those of shapes or traditional graphical elements—in particular, terms that capture the observed similarity and repetition alone. Computationally, determining the fractal representation of an image requires the use of the fractal encoding algorithm, which, given an image D , seeks to discover the set of transformations T that can transform any source image into D .

Decompose D into a set of N smaller images $\{d_1, d_2, d_3, \dots, d_n\}$. These individual images are sets of points.

For each image d_i :

Examine the entire source image S for an equivalent image s_i such that a similitude transformation of s_i will result in d_i . This transformation will be a 3×3 matrix, as the points within s_i and d_i under consideration can be represented as the 3D vector $\langle x, y, c \rangle$ where c is the (grayscale) color of the 2D point $\langle x, y \rangle$.

Collect all such transforms into a set of candidates C .

Select from C the transform which most minimally achieves its work, according to some predetermined, consistent metric.

Let T_i be the representation of the chosen affine transformation of s_i into d_i .

The set $T = \{T_1, T_2, T_3, \dots, T_n\}$ is the fractal encoding of the image D .

Algorithm 1. Fractal encoding algorithm for determining the fractal representation of an image D .

This algorithm, shown in Algorithm 2, is considered “fractal” for two reasons: first, the transformations chosen are generally contractive, which leads to convergence, and second, the convergence of S into D can be shown to be the mathematical equivalent of considering D to be an attractor (Barnsley & Hurd, 1992).

Once fractal representations have been calculated for each pair of images in a Raven's problem, the metric shown in Equation (2) is used to calculate similarity between all of the pairwise relationships present in the matrix and those calculated with the given answer choices, using features derived from the fractal encodings. Whichever answer choice yields the most similar fractal representations across all pairwise relationships is chosen as the final answer. The fractal method is described in more detail in McGreggor, Kunda, and Goel (2010).

Results

We tested both the affine and fractal algorithms on all 60 problems from the Raven's Standard Progressive Matrices (SPM) test. To obtain visual inputs for the algorithms, we first scanned a paper copy of the SPM, aligned each page to lie squarely along horizontal and vertical axes, and then divided each problem into separate image files representing each of the matrix entries and answer choices. No further image processing was performed on these input images. As a result, these images were fairly noisy; they contained numerous misalignments and pixel-level artifacts from the scanning and subdividing processes.

Then, after answers for all 60 SPM problems were obtained from each algorithm, we scored each method according to standard protocols for the SPM. In particular, we looked at three different measures of performance:

- 1) The total score from the SPM summarizes the test-taker's overall level of performance.
- 2) This total score can be compared to national age-group norms to determine a percentile ranking.
- 3) A “consistency” measure is obtained by comparing performance on each of the five sets within the SPM, A through E, with the expected scores for each set given the same total score, which are obtained from normative data (Raven, Raven, & Court, 1998).

In addition, we conducted a separate analysis of results according to problem type, looking at accuracy as a function of three problems classifications: “gestalt continuation,” “visuospatial,” and “verbal-analytic,” which we obtained from a published factor analytic study of the SPM (Lynn, Allik, & Irving, 2004).

Affine Results

The affine algorithm correctly solved 35 of the 60 problems on the SPM. For children in the U.S., this total score corresponds to the 75th percentile for 9-year-olds, the 50th percentile for 10½-year-olds, and the 25th percentile for 13-year-olds (Raven, Raven, & Court 1998).

The breakdown of this total score across sets is shown in Figure 2, along with the expected score composition for this

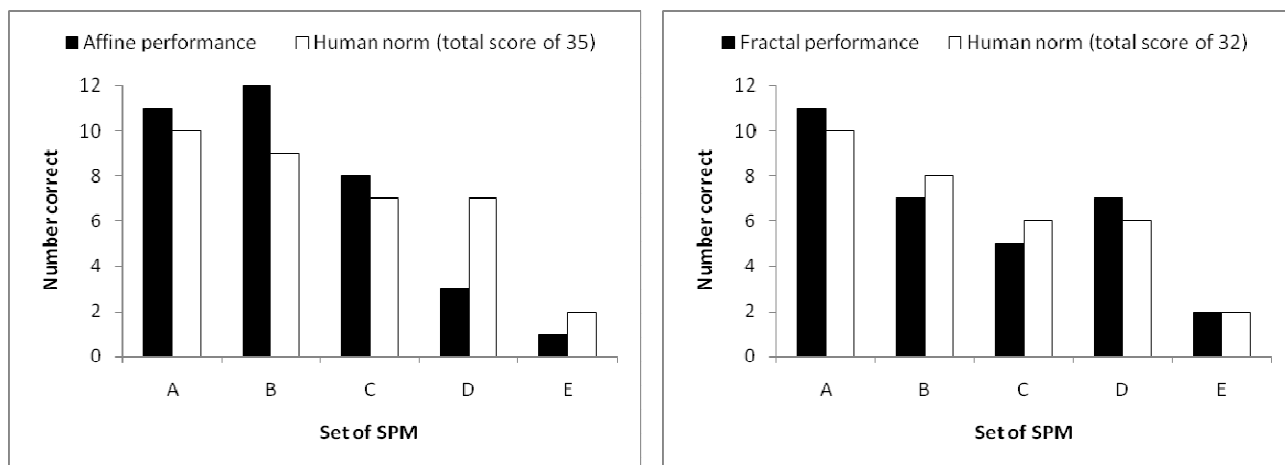


Figure 2: Breakdown of affine (left) and fractal (right) results across sets in the SPM. Also shown is the expected score breakdown for total scores of 35 and 32, from normative human data (Raven, Raven, & Court, 1998).

same total score. Scoring instructions for the SPM indicate that, if the score for any set deviates from the expected score for that set by more than two, the overall test results cannot necessarily be interpreted as a measure of general cognitive function (Raven, Raven, & Court, 1998). This check is intended to detect scores affected by a poor understanding of test instructions, random guessing strategies, or other departures from the intended test-taking framework. As shown in Figure 2, the affine scores deviate by more than ± 2 from the expected scores on sets B and D. In particular, the affine algorithm does too well on Set B and not well enough on Set D to match typical human norms.

Fractal Results

The fractal algorithm correctly solved 32 of the 60 problems on the SPM. For children in the U.S., this total score

corresponds to the 75th percentile for 8-year-olds, the 50th percentile for 9½-year-olds, and the 25th percentile for 11½-year-olds (Raven, Raven, & Court 1998).

The breakdown of this total score across sets is shown in Figure 2, along with the expected score composition for this same total score. The fractal scores fall within ± 2 of the expected scores for each set, indicating that the fractal results are “consistent” with normative SPM scores.

Results by Problem Type

The final analysis we performed looked at the performance of both algorithms as a function of problem type on the SPM. Factor-analytic studies have often found evidence for multiple factors underlying problem solving on the SPM (e.g. van Der Ven & Ellis, 2000); we used the breakdown obtained by one such study to divide problems into those

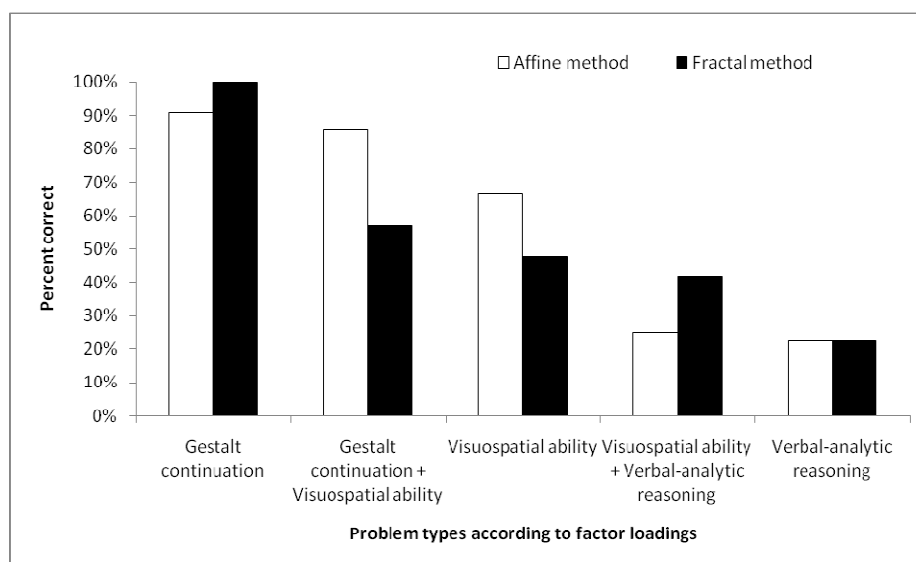


Figure 3: Breakdown of affine and fractal algorithm results on the SPM by problem type. Problem breakdowns were obtained from a factor-analytic study of human performance (Lynn, Allik, & Irving, 2004).

that loaded on “gestalt continuation,” “visuospatial,” or “verbal-analytic” factors (Lynn, Allik, & Irving, 2004).

Figure 3 shows the performance of both the affine and fractal algorithms on problems from the SPM which load on different combinations of these factors. Both the affine and fractal methods perform most strongly on gestalt problems, slightly less so visuospatial problems, and significantly less so on problems requiring verbal-analytic reasoning, though the relative difficulties of each of these problem types could represent a potential confound for these results.

Discussion

We have presented two different algorithms that use purely visual representations and transformations to solve more than half of the problems on the Raven’s SPM test. Our results align strongly with evidence from typical human behavior suggesting that multiple cognitive factors underlie problem solving on the SPM, and in particular, that some of these factors appear based on visual operations. Whether these algorithms behave on the SPM similarly to individuals with autism, who may demonstrate a cognitive preference for solving the test visually, remains to be determined.

That purely visual methods can achieve such significant results on a standardized intelligence test is a little surprising to us, especially as the input images for both algorithms were taken “as is,” from raw scans of a paper copy of the test. This robust level of performance calls attention to the visual processing substrate shared by the affine and fractal algorithms: similitude transforms as a mechanism for image manipulation, and the ratio model of similarity as a mechanism for image comparison. Of course, there are many other types of visual processing that may or may not be important for accounts of visual analogy, such as non-similitude shape transformations or image convolutions, which certainly bear further investigation.

While it has been shown (Davies, Yaner, & Goel, 2008) that visuospatial knowledge alone may be sufficient for addressing many analogy problems, the representations used in that work were still propositional. In contrast, the methods described here use only visual representations in the form of image similitude transformations. We believe the visual methods we have presented for solving the SPM can be generalized to visual analogy in other domains, such as other standardized tests (e.g. the Miller’s Geometric Analogies test). We conjecture that these methods may provide insight into general visual recognition and recall.

Acknowledgments

This research has been supported by an NSF grant (IIS Award #0534266), “Multimodal Case-Based Reasoning in Modeling and Design,” by ONR through an NDSEG fellowship, and by the NSF GRFP fellowship program.

References

Barnsley, M. F., & Hurd, L. P. (1992). *Fractal Image Compression*. Boston, MA: A.K. Peters.

- Bringsjord, S., & Schimanski, B. (2003). What is artificial intelligence? Psychometric AI as an answer. *IJCAI*, 18, 887–893.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404–31.
- Davies, J., Goel, A., & Yaner, P. (2008). Proteus: A theory of visual analogies in problem solving. *Knowledge-Based Systems*, 21(7), 636–654.
- Dawson, M., Soulières, I., Gernsbacher, M. A., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18(8), 657–662.
- DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven’s advanced progressive matrices: Evidence for multidimensional performance determinants. *Intelligence*, 21(2), 135–155.
- Dillon, R. F., Pohlmann, J. T., & Lohman, D. F. (1981). A factor analysis of Raven’s Advanced Progressive Matrices freed of difficulty factors. *Educational and Psychological Measurement*, 41, 1295–1302.
- Hunt, E. (1974). Quote the raven? Nevermore! In L. W. Gregg (Ed.), *Knowledge and Cognition* (pp. 129–158). Hillsdale, NJ: Erlbaum.
- Kunda, M., & Goel, A. K. (2008). Thinking in Pictures: A Fresh Look at Cognition in Autism. In *Proc. 30th Annual Conf. Cognitive Science Society* (pp. 321–326).
- Lovett, A., Forbus, K., & Usher, J. (2007). Analogy with qualitative spatial representations can simulate solving Raven’s Progressive Matrices. In *Proc. 29th Annual Conf. Cognitive Science Society* (pp. 449–454).
- Lynn, R., Allik, J., & Irving, P. (2004). Sex differences on three factors identified in Raven’s Standard Progressive Matrices. *Intelligence*, 32(4), 411–424.
- Mackintosh, N., & Bennett, E. (2005). What do Raven’s Matrices measure? An analysis in terms of sex differences. *Intelligence*, 33(6), 663–674.
- McGreggor, K., Kunda, M., & Goel, A. (2010). A fractal approach towards visual analogy. In *Proc. 1st International Conf. Computational Creativity*, Lisbon, Portugal, January, 2010.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven’s Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.
- Soulières, I., Dawson, M., Samson, F., Barbeau, E. B., Sahyoun, C. P., Strangman, G. E., et al. (2009). Enhanced visual processing contributes to matrix reasoning in autism. *Human Brain Mapping*, 30(12), 4082–107.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.
- van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven’s standard progressive matrices. *Personality and Individual Differences*, 29(1), 45–64.
- Vigneau, F., & Bors, D. A. (2008). The quest for item types based on information processing: An analysis of Raven’s Advanced Progressive Matrices, with a consideration of gender differences. *Intelligence*, 36(6), 702–710.