Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence

Two Visual Strategies for Solving the Raven's Progressive Matrices Intelligence Test

Maithilee Kunda, Keith McGregor, and Ashok Goel

School of Interactive Computing, Georgia Institute of Technology
{mkunda, keith.mcgregor}@gatech.edu, goel@cc.gatech.edu

Abstract

We present two visual algorithms, called the affine and fractal methods, which each solve a considerable portion of the Raven's Progressive Matrices (RPM) test. The RPM is considered to be one of the premier psychometric measures of general intelligence. Current computational accounts of the RPM assume that visual test inputs are translated into propositional representations before further reasoning takes place. We propose that visual strategies can also solve RPM problems, in line with behavioral evidence showing that humans do use visual strategies to some extent on the RPM. Our two visual methods currently solve RPM problems at the level of typical 9- to 10-year-olds.

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

Introduction

In previous work (Kunda, McGreggor, & Goel, 2010), we presented two visual algorithms for solving problems from Raven's Progressive Matrices (RPM) intelligence tests, along with experimental results. Here, we summarize this recent work and discuss its implications for AI.

The RPM consists of geometric analogy problems, as shown in Figure 1; problems contain either 2x2 (like the example given in Figure 1) or more complex 3x3 matrices. Although the RPM is supposed to measure only eductive ability (Raven, Raven, & Court, 2003), its high correlation with other IQ tests has rendered it a premier psychometric measure of general intelligence, and it is widely used in clinical, educational, occupational, and scientific settings.

Computational accounts of problem solving on the RPM have generally assumed that visual inputs are translated into propositions before further reasoning takes place. Carpenter et al. (1990) implemented a production system that took hand-coded propositional descriptions of RPM inputs and then chose from predefined rules to solve each problem. Bringsjord and Schimanski (2003) used a theorem-prover to solve RPM problems stated in first-order logic. Lovett, Forbus, & Usher (2007) combined automated sketch understanding with structure-mapping to solve RPM problems, which were represented using hand-drawn vector graphics. Cirillo and Strom (2010) also used vector graphics representations of RPM problems along with a set of pre-defined patterns to predict an answer. Finally, Rasmussen & Eliasmith (2011) used a spiking neuron model to induce rules for solving RPM problems represented as hand-coded propositional vectors.

Despite considerable differences in architecture and focus, all of these computational approaches have been similar in two respects. First, as mentioned earlier, they all rely on some translation process to convert visual test inputs into propositional representations, which then become the sole form of representation used by each system. Second, each system posits only one fundamental problem-solving strategy; individual differences in RPM performance are assumed (either implicitly or explicitly) to stem from quantitative variations in this strategy, rather than from qualitative differences in the strategy itself.

In contrast, Hunt (1974) proposed two qualitatively different strategies: Gestalt, using visual representations and perceptual operations, and Analytic, using propositions and logical operations. While neither RPM algorithm was

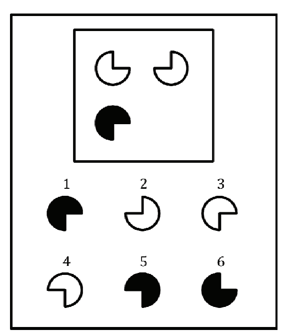

Figure 1: Example problem similar to one from the Standard Progressive Matrices (SPM) test. The correct answer is 5.

For each base transform T:
Apply T to Image A.
Find best-match translation (tx, ty) 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 max. is (1), then X = 0.
If max. is (2), then X = B – A, and $\oplus$ is image addition.
If max. is (3), then X = A – B, and $\oplus$ is image subtraction.
The best-fit similitude transformation can then be specified as:
$$[T_{max}+(tx, ty)](A) \oplus X = B$$

Decompose D into a set of N smaller images $\{d_1, d_2, d_3, \ldots, d_n\}$. 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 3x3 matrix, as the points within s_i and d_i under consideration can be represented as the 3D vector <x, y, c> where c is the (grayscale) color of the 2D point <x,y>.
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 represent the chosen affine transformation of s_i into d_i.
The set $T = \{T_1, T_2, T_3, \ldots, T_n\}$ is the fractal encoding of the image D.

Figure 2. Algorithm 1 (left): Affine method. Algorithm 2 (right): Fractal method.

implemented, a theoretical analysis suggested that both methods would be equally effective for certain problems.

In addition to Hunt's computational argument, there is evidence that humans use qualitatively different RPM strategies. Within-individual strategy differences have been studied as a function of problem type, primarily through factor analyses (Lynn et al., 2004), often dividing problems into those solvable with visual versus verbal operations.

Between-individual strategy differences have emerged in studies of autism. Whereas the RPM scores of typically developing (TD) individuals are highly correlated with full IQ scores, individuals with autism often show much higher RPM scores (Dawson et al., 2007), possibly because intact visual abilities in autism can be recruited to solve many RPM problems (Kunda & Goel, in press). Recent fMRI data of the RPM showed that individuals with autism had lower brain activation in areas associated with language and working memory and higher activation in visual areas than did TD individuals (Soulières et al., 2009).

Visual Methods for the Raven's Test

We have developed two different RPM algorithms that we call the "affine" method and the "fractal" method. Unlike previous computational models of the RPM, these methods use image transformations to solve RPM problems directly, without first converting visual inputs into propositions. We define a representation as being purely visual if it captures image information only at the pixel level, i.e. as a spatial array of individual color/intensity values. A representation is propositional if it encodes higher-level visual entities as propositions, e.g. as labeled lines, shapes, textures, etc.

Figure 2 outlines the core mechanisms for both visual algorithms. Due to space constraints, we present only brief descriptions of these algorithms below; more details can be found in (Kunda, McGreggor, & Goel, 2010).

At the core of these methods are affine transformations, and in particular *similitude transforms*, which can be represented as compositions of dilation, orthonormal transformation, and translation. We presently use the identity transform, horizontal and vertical reflections, and 90°, 180°, and 270° rotations, composed with translation. 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 (Kosslyn, Thompson, & Ganis, 2006).

Similarity also lies at the core of both methods, as calculated using the *ratio model* (Tversky, 1977):

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

In Eq. (1), f represents some function over features in each of the specified sets, and α and β are 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:

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

Eq. (2) is used in both methods, and it yields maximal similarity when A is equal to B. In contrast, if α is one and β is zero, it yields maximal similarity when A is a proper subset of B. If α is zero and β is one, then maximal similarity is found when B is a proper subset of A.

The Affine Method

The affine method assumes that 1) elements within a row or column in an RPM problem matrix are related by similitude transforms, and 2) analogical relationships exist, in the form of identical similitude transforms, across parallel rows or columns of the matrix. Each similitude transform is represented as a combination of three image operations: base transform, translation, and composition.

First, the algorithm determines which transform best fits any complete row or column in the matrix; Algorithm 1 in Figure 2 shows how, for images A and B, the "best-fit" transform is found. Then, this transform is applied to whichever parallel row/column contains the missing entry to generate a guess image for the answer. Finally, this guess is compared to each answer choice, using Eq. (2), and the best match is chosen as the answer.

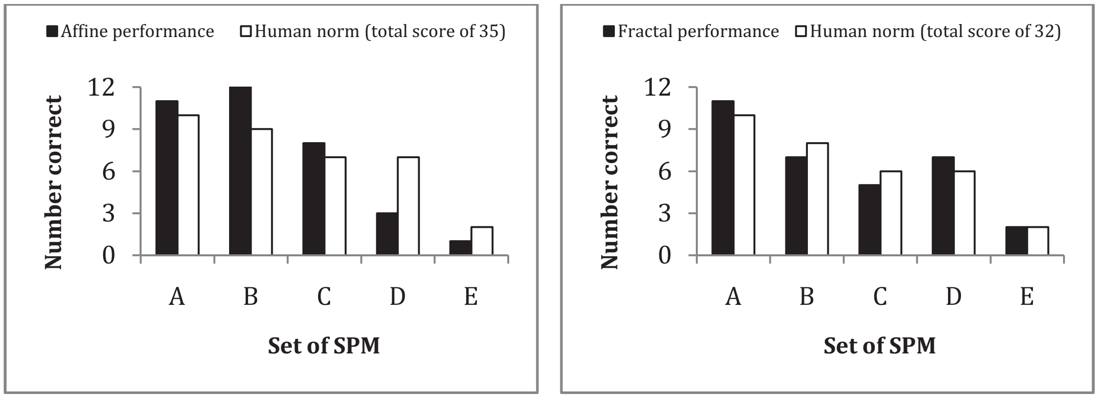

Figure 3: Breakdown of affine (left) and fractal (right) results across sets in the SPM. Also shown is the expected score breakdown for total scores of 35 and 32, from normative human data (Raven, Raven, & Court, 2003).

The Fractal Method

Like the affine method, the fractal method seeks to find a re-representation of the images within an RPM problem as a set of similitude transforms. Unlike the affine method, the fractal method seeks representations at a significantly finer partitioning of the images, and uses features derived from these resulting "fractal" 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 on the notion of real world images, i.e. images that are two dimensional and continuous (Barnsley & Hurd, 1992), and draws upon the 1) repetition and 2) similarity at different scales that are found in such images. Fractal representations seek to describe images in terms other than those of shapes or traditional graphical elements—i.e. terms that capture this observed similarity and repetition alone. Computationally, determining fractal representations uses the fractal encoding algorithm, shown in Algorithm 2 in Figure 2.

Once fractal representations have been calculated for each pair of images in the problem matrix, Eq. (2) is used to calculate similarity between all pairwise relationships 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 (McGreggor, Kunda, & Goel, 2010).

Results

We tested both the affine and fractal methods on the complete Standard Progressive Matrices (SPM) test. The SPM consists of 60 problems divided into five sets labeled A-E. To obtain visual inputs for the algorithms, we first scanned a paper copy of the SPM, aligned each page squarely, and then divided each problem into separate image files for each matrix entry and answer. No further image processing was performed on these images.

The affine algorithm correctly solved 35 of 60 problems on the SPM, as shown in Figure 3. For children in the U.S., this total score corresponds to the 50th percentile for 10½-year-olds (Raven, Raven, & Court, 2003). The fractal algorithm correctly solved 32 of 60 problems, which corresponds to the 50th percentile for 9½-year-olds.

We also looked at the performance of both methods as a function of problem type on the SPM; we used the results from a factor analysis to divide problems into those loading on "gestalt continuation," "visuospatial," or "verbal-analytic" factors (Lynn, Allik, & Irving, 2004). Figure 4 shows the performance of both methods with respect to problem type. 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 problem difficulty represents a potential confound.

Discussion

We presented results from (Kunda, McGreggor, & Goel, 2010) describing two algorithms that use purely visual representations to solve more than half of the problems on the Raven's SPM test. These results are significant not only because we have shown purely visual methods to be surprisingly powerful in addressing a standard test of general intelligence, which is itself an accomplishment according to the psychometric AI school of thought (Brinsjord & Schimanski, 2003), but also because they illustrate the capabilities of visual methods in general.

As previous RPM models have shown, as well as other visual analogy work in AI (Davies, Yaner, & Goel, 2008), visuospatial knowledge alone, represented propositionally, can be sufficient to solve certain analogy problems. In this work, we have shown that it is not necessary to translate

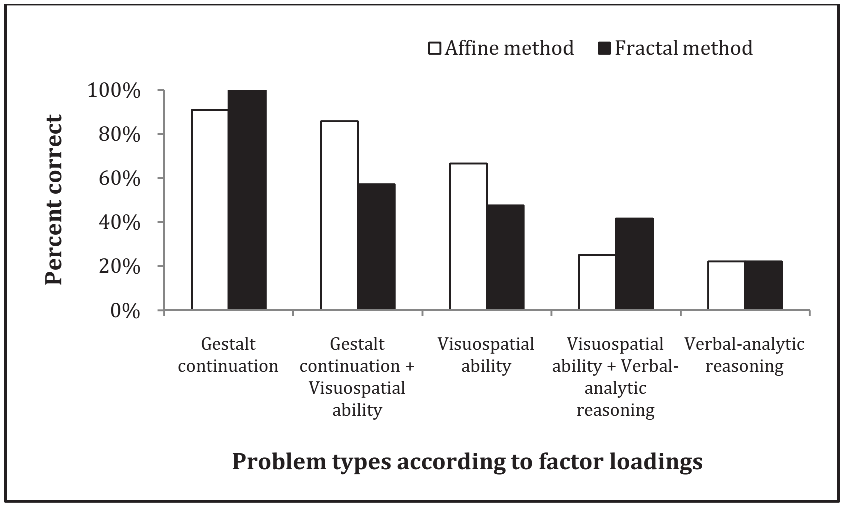

Figure 4: Breakdown of affine and fractal algorithm results on the SPM by problem type. Problem breakdowns were obtained from a factor-analytic study of human performance (Lynn, Allik, & Irving, 2004).

images into propositions at all for certain reasoning to take place. The analogical (i.e. having a structure corresponding to what is represented) properties of real-world images, including their amenability to affine transforms, repetition, and self-similarity, appear sufficient to support higher-level reasoning, and the ratio model of similarity is one instantiation of a comparison mechanism that does not rely on the extraction of propositional features.

In general, the results of this work support the view of perception as a first-class object for reasoning in AI systems. Many open questions remain for the study of visual strategies on the RPM, such as what role other visual operations like convolution or deformation might play, and how visual and propositional strategies might be seamlessly bridged in a single cognitive agent.

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., Just, M., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Rev.*, 97, 404-31.

Cirillo, S., & Strom, V. (2010). An anthropomorphic solver for Raven's Progressive Matrices (No. 2010:096). Goteborg, Sweden: Chalmers University of Technology.

Davies, J., Goel, A., & Yaner. P. (2008). Proteus: Visuospatial analogy in problem solving. *Knowledge-Based Systems*, 21, 636-654.

Dawson, M., Soulières, I., Gernsbacher, M., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18, 657-662.

Hunt, E. (1974). Quote the raven? Nevermore! In L. Gregg (Ed.), *Knowledge and Cognition* (pp. 129–158). Hillsdale, NJ: Erlbaum.

Kunda, M., & Goel, A. (in press). Thinking in Pictures as a cognitive account of autism. *J. Autism and Dev. Disorders*.

Kunda, M., McGreggor, K., & Goel, A. (2010). Taking a look (literally!) at the Raven's intelligence test: Two visual solution strategies. In *Proc. 32nd Conf. Cog. Sci. Society* (pp. 1691-1696)

Lovett, A., Forbus, K., & Usher, J. (2007). Analogy with qualitative spatial representations can simulate solving Raven's Progressive Matrices. In *Proc. 29th Conf. Cog. Sci. Society* (pp. 449-454).

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's SPM. *Intelligence*, 32, 411-424.

McGreggor, K., Kunda, M., & Goel, A. (2010). A fractal approach towards visual analogy. In *Proc. 1st International Conf. Computational Creativity*.

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

Raven, J., Raven, J. C., & Court, J. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Soulières, I., Dawson, M., Samson, F., Barbeau, E., Sahyoun, C., Strangman, G., Zeffiro, T., & Mottron, L. (2009). Enhanced visual processing contributes to matrix reasoning in autism. *Human Brain Mapping*. 30, 4082-107.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.