Available online at www.sciencedirect.com

SciVerse ScienceDirect

Cognitive Systems Research 22–23 (2013) 47–66

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations

Action editor: Paul Thagard

Maithilee Kunda *, Keith McGreggor, Ashok K. Goel

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

Received 30 May 2012; accepted 6 July 2012
Available online 11 August 2012

Abstract

We describe a computational model for solving problems from Raven's Progressive Matrices (RPM), a family of standardized intelligence tests. Existing computational models for solving RPM problems generally reason over amodal propositional representations of test inputs. However, there is considerable evidence that humans can also apply imagery-based reasoning strategies to RPM problems, in which processes rooted in perception operate over modal representations of test inputs. In this paper, we present the "affine model," a computational model that simulates modal reasoning by using iconic visual representations together with affine and set transformations over these representations to solve a given RPM problem. Various configurations of the affine model successfully solve between 33 and 38 of the 60 problems on the Standard Progressive Matrices, which matches levels of performance for typically developing 9- to 11-year-old children. This suggests that, for at least a sizeable subset of RPM problems, it is not always necessary to extract amodal symbols in order to arrive at the correct answer, and iconic visual representations constitute a sufficient form of representation to successfully solve these problems. We intend for the affine model to serve as a complementary computational account to existing propositional models, which together may provide an integrated, dual-process account of human problem solving on the RPM.

Keywords: Affine transformations; Analogy; Iconic representations; Intelligence tests; Mental imagery; Raven's Progressive Matrices

1. Introduction

Raven's Progressive Matrices (RPM) is a collection of widely-used standardized intelligence tests consisting of 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. Figs. 1 and 2 show examples of two-by-two (2×2) and three-by-three (3×3) matrix problems, respectively, which are similar to actual RPM problems.[1]

There are currently three published versions of the RPM: (1) the original Standard Progressive Matrices (SPM), (2) the Advanced Progressive Matrices (APM), developed as a more difficult test to reduce the ceiling effects sometimes found with the SPM, and (3) the Colored Progressive Matrices (CPM), intended as a simpler test than the SPM to be used with children, the elderly, or other individuals falling into lower IQ brackets (Raven, Raven, & Court, 2003). For the remainder of this paper, we use the term RPM to refer to the above family of tests, and we use the labels SPM, APM, and CPM to refer to specific members of the test family.

* Corresponding author.
E-mail addresses: mkunda@gatech.edu (M. Kunda), keith.mcgreggor@gatech.edu (K. McGreggor), goel@cc.gatech.edu (A.K. Goel).

[1] To protect the confidentiality of the RPM, we present example problems that are similar, but not identical, to actual test problems.

1389-0417/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cogsys.2012.08.001

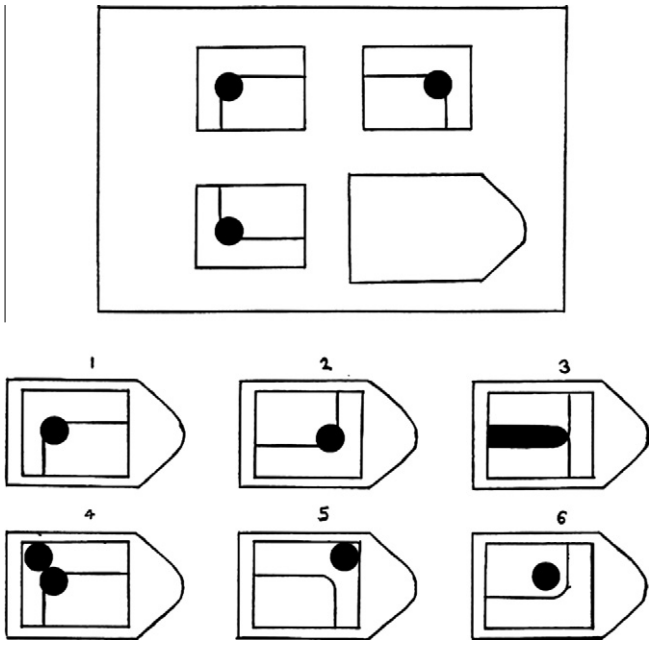


Fig. 1. 2×2 example problem similar to those from the Raven's Progressive Matrices family of tests.

The RPM tests were originally designed to measure only eductive ability, or the ability to extract and understand information from a complex situation, which is sometimes referred to as "fluid intelligence" (Raven et al., 2003). They were intended to be used together with the Mill Hill Vocabulary Scales, which measure reproductive ability, or the ability to recall previously learned information, sometimes called "crystallized intelligence." Together, these two tests would provide a measure of Spearman's general intelligence factor *g*, which Spearman had supposed could be decomposed into eductive and reproductive components (Spearman, 1923). However, over time, it was found that the RPM alone exhibited a very high level of correlation with other intelligence tests, leading the RPM to become widely considered one of the best single psychometric measures of *g* (Snow, Kyllonen, & Marshalek, 1984).

Using the RPM as a measure of general intelligence, though it consists only of problems in a single, visual format, stands in contrast to using broader IQ tests like the Wechsler scales, which are comprised of subtests that span several different verbal and nonverbal domains. In fact, the RPM was originally developed as an easy-to-administer, easy-to-score alternative to traditional multi-domain intelligence tests, which can take many hours to administer and often yield complex, multi-dimensional subscores which must then be combined to create a final IQ score (Raven et al., 2003). Due to its ease of administration and scoring, as well as the fact that it requires little verbal instruction or explicit verbal comprehension, the RPM is widely used as a test of general intelligence in clinical, educational, occupational, and scientific settings.

1.1. Computational models of problem-solving on the RPM

Computational accounts of problem solving must specify what kinds of representations are used to contain problem information and what types of processes operate over these representations to generate solutions. Following

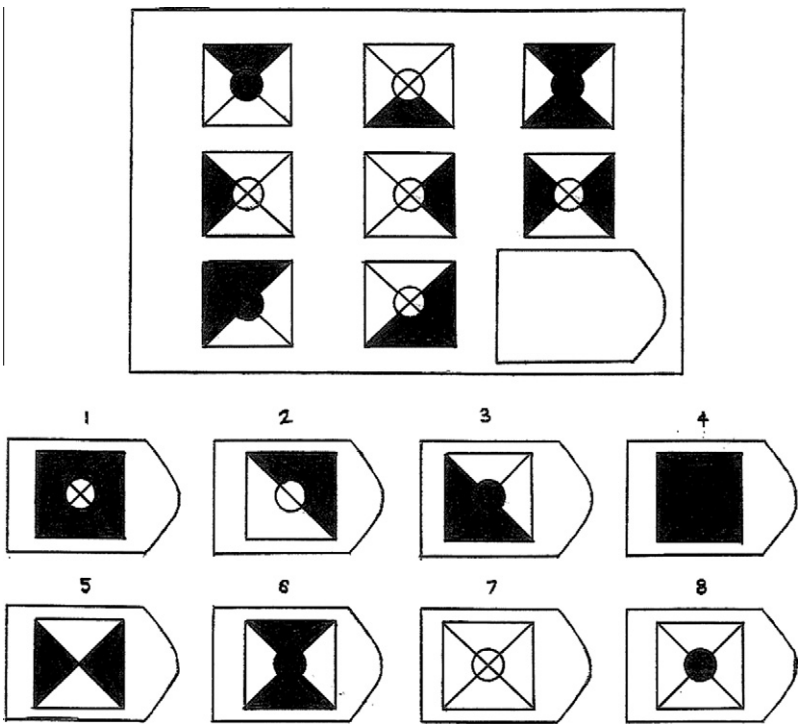


Fig. 2. 3×3 example problem similar to those from the Raven's Progressive Matrices family of tests.

Nersessian (2008), we categorize representations of RPM problems along two dimensions:

(a) *Iconic versus propositional*: The term *iconic* refers to representations that are analogical, in the sense that they carry some structural correspondence to what they represent. In the case of iconic visual representations, this structural correspondence takes the form of shared spatial relationships. *Propositional* representations, on the other hand, carry no such correspondence between format and content.
(b) *Modal versus amodal*: Symbols used in an iconic representation can be either *modal* or *amodal*, depending on whether they are rooted in perceptual states. (Symbols used in a propositional representation are, by their nature, amodal.)

A linguistic description of the shapes and relations in an RPM problem would constitute an amodal propositional representation, for example: **is-left-of(triangle, circle)**. A diagram indicating the spatial layout of shapes with each shape described linguistically would constitute an amodal iconic representation, in that the representation does show some structural correspondence with the problem, but the linguistic symbols are not themselves directly related to perceptual inputs, for example: **triangle–circle**. An image showing both the spatial layout of shapes as well as their visual appearance would constitute a modal iconic representation, in that the representation shows structural correspondence with the problem as well as with the visual perceptual state that might be generated by looking at it, for example: ▲–●.

The main representational distinction on which we focus is between computational models that use *amodal propositional representations* of RPM problems and computational models that use *modal iconic representations*.[2] We discuss in the next section how these representational types might correspond to qualitatively different problem-solving strategies used by humans when solving RPM problems.

Hunt (1974) proposed the existence of two qualitatively different RPM strategies which varied primarily in how problem inputs were represented. Hunt's "Analytic" algorithm used amodal propositional representations and operations such as constancy and addition/subtraction. This algorithm proceeded by first abstracting features from each matrix entry and then iteratively applying operators to the entries within a row or column, or to an entire row or column, to generate partial answer predictions. If the predicted answer was found among the answer choices and was unique, then the algorithm halted. If either of these conditions were not met, then the algorithm iterated to predict a different answer or to refine the current partial answer.

[2] To our knowledge, there have been no computational models of the RPM which use amodal iconic representations.

Hunt's "Gestalt" algorithm, akin to mental imagery, used modal iconic representations and perceptual operations like continuation and superposition to solve RPM problems. The algorithm successively applied various visual operations to entries from the matrix in order to obtain an answer that matched one given in the answer choices, using an answer-iteration procedure similar to that used in the Analytic algorithm. However, neither algorithm was actually implemented.

All of the RPM models that have since been developed resemble Hunt's Analytic algorithm in that they rely on a conversion of problem inputs into amodal propositional representations. None of these models have adopted the approach suggested by Hunt in his Gestalt algorithm.

Model #1: Carpenter, Just, and Shell (1990) implemented a production system that took as input hand-coded propositional descriptions of problems from the APM. The system chose from a predefined set of rules over matrix elements in order to predict an answer for each problem. The predicted answer was compared to the answer choices in order to choose the best match. The predefined rules were generated by the authors from an *a priori* inspection of the APM and were validated in experimental studies by observing what "rules" participants used while taking the test, as evidenced by verbal reporting protocols. Differences between low- and high-scoring participants were modeled by developing two different versions of the production system; the more advanced system contained an increased vocabulary of rules and a goal monitor for setting and adjusting the high-level problem-solving process being used. Both systems were tested against 34 of the 48 problems from the APM. The basic system solved 23 of these 34 problems, while the more advanced system solved 32 of the 34 problems.

Model #2: Bringsjord and Schimanski (2003) used a theorem-prover to solve selected RPM problems stated in first-order logic, though no specific results were reported.

Model #3: Lovett, Forbus, and Usher (2010) combined automated sketch understanding with the structure-mapping technique for analogy to solve problems from the SPM. Input images from the test were first redrawn by hand in Powerpoint, and the resulting vector graphics objects were fed into the system. The system translated these inputs into amodal propositional descriptions using a procedure for automated sketch understanding. Then, a series of strategies based on the structure-mapping technique for analogy were applied to detect certain patterns of structural relationships between various elements in the matrix. These derived structural relationships were also used to refine object segmentation and groupings by revisiting the original vector-graphics-based representations and extracting modified propositional descriptions that allowed for improved structural matches. Finally, each answer choice was inserted into the matrix, and the answer providing the closest matching structural relationship within the matrix was selected. This system was tested against 48 of the 60 problems on the SPM and solved 44 of these 48 problems.

Model #4: The system of Cirillo and Ström (2010), like that of Lovett et al. (2010), took as inputs hand-drawn vector graphics representations of test problems and used an automated procedure to create hierarchical propositional representations of the problem information. Then, like the work of Carpenter et al. (1990), the system drew from a set of pre-defined patterns, derived by the authors from an *a priori* inspection of the SPM, to find the best-fit pattern for a given problem. The resulting pattern was used to predict an answer, though no explicit procedure was given for matching the predicted answer to one of the given answer choices. This system was tested against 36 of the 60 problems from the SPM and solved 28 of these 36 problems.

Model #5: Rasmussen and Eliasmith (2011) used a spiking neuron model to induce rules for solving RPM problems. Input images from the test were first hand-coded into vectors of propositional attribute-value pairs, and then the spiking neuron model was used to derive several individual transformations among these vectors and abstract over them to induce a general rule transformation for that particular problem. While the authors attested that this system could correctly solve RPM problems, they did not present any results regarding which specific tests or problems were addressed.

As mentioned above, despite considerable differences in computational architecture and in problem-solving focus, all five computational models of the RPM that have actually been implemented and tested have reasoned over amodal propositional representations of test inputs. In addition, each model posits only one fundamental problem-solving strategy; individual differences in human RPM performance are assumed to stem solely from quantitative variations in this core strategy, rather than (as Hunt suggested might be possible) from qualitative differences between distinct strategies.

1.2. Human problem-solving on the RPM

There is considerable evidence that humans recruit qualitatively different strategies on the RPM in terms of what types of mental representations and operations are used for reasoning. In particular, the main contrast observed in the literature on human RPM problem solving is between *visual* and *verbal* strategies. Visual strategies are those that use modal iconic mental representations rooted in the visual perceptual modality; the use of mental imagery would fall into this category. Verbal strategies use amodal propositional mental representations, such as linguistic descriptions of RPM problems.

One way in which strategy differences have been studied is as a function of problem type on the RPM tests, primarily through factor analyses of the SPM (Lynn, Allik, & Irwing, 2004; van der Ven & Ellis, 2000) and of the APM (Dillon, Pohlmann, & Lohman, 1981; Mackintosh & Bennett, 2005; Vigneau & Bors, 2008). These studies have identified multiple factors underlying RPM tests, indicating variations in the recruitment of particular cognitive mechanisms for different problems, and have often divided test problems into two primary categories: those solved using visuospatial or gestalt reasoning and those solved using verbal or analytic reasoning (though it should be pointed out that, while the factor loadings themselves are statistically determined, labels for the various factors appear to be based on the authors' own inspections of problem groupings by factor). Following the Gestalt/Analytic strategy divide proposed by Hunt (1974), Kirby and Lawson (1983) studied the performance effects of training students to use a particular strategy; part of this study involved developing a new series of test items on which the type of strategy being used led to a different selection of a "correct" answer choice, thus demonstrating the existence of strategy-linked answer types (in addition to strategy-linked problem types).

From neuroscience, one fMRI study of RPM performance (Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997) found that patterns of brain activity differed significantly based on whether participants were solving "figural" versus "analytic" problems, using problem classifications derived from the Carpenter et al. (1990) computational work. Figural problems were found to induce brain activity primarily in spatial and object working memory regions, while analytic problems induced additional brain activity in verbal working memory and executive processing regions. Studies of patients with focal brain lesions have also found linkages between brain regions associated with specific types of visual or verbal processing and successful performance on figural versus analytic problems (Berker & Smith, 1988; Villardita, 1985).

DeShon, Chan, and Weissbein (1995) had participants complete the APM while simultaneously performing a "verbal overshadowing" protocol, in which they had to verbally describe their reasoning. The authors hypothesized that requiring overt verbal descriptions would bias participants towards using verbal instead of visual strategies and thereby impair performance on problems that would normally have been solved visually, a pattern which was borne out in the resulting data. These findings also call into question the experimental methodology of using verbal reporting protocols as a window into RPM problem solving, as the act of verbal reporting itself may cause shifts in an individual's strategy; this "verbal overshadowing" phenomenon has been observed in other problem domains as well (Schooler & Engstler-Schooler, 1990; Schooler, Ohlsson, & Brooks, 1993).

Differences in strategy on the RPM have also recently come to light in studies of individuals with autism and their RPM performance in comparison with typically developing (TD) individuals. Evidence across many task domains suggests that individuals with autism may exhibit a general bias towards using visual mental representations over verbal ones (Kunda & Goel, 2008, 2011), and on the RPM in particular, Soulières et al. (2009) found using fMRI that individuals with autism had lower brain activation in

prefrontal and parietal areas associated with language and working memory and higher activation in visual occipital areas than did TD individuals. On a related but non-RPM set of matrix reasoning tasks, Sahyoun, Soulières, Belliveau, Mottron, and Mody (2009) found through examinations of response latency that individuals with autism exhibited a bias towards using primarily visuospatial mediation, whereas TD individuals and individuals with Asperger's seemed able to additionally recruit verbal mediation in solving the problems. Finally, whereas, as mentioned earlier, the RPM scores of TD individuals are generally very well matched with their full IQ scores such as from the Wechsler scales, individuals with autism have often demonstrated RPM scores much higher than their Wechsler scores (Bölte, Dziobek, & Poustka, 2009; Dawson, Soulières, Gernsbacher, & Mottron, 2007), which is consistent with the notion of a reliance on visual strategies that might be sufficient for solving visually presented RPM problems but not for completing a full, multi-domain IQ test.

1.3. Motivation

In summary, there is considerable evidence from both behavioral and neuroimaging studies that humans use qualitatively different cognitive strategies to solve RPM problems, in terms of what types of mental representations and operations are employed. Some strategies appear to be based around modal iconic representations and cause neural activity in brain regions associated with visual and spatial processing, whereas other strategies appear to be based on amodal propositional representations and cause neural activity in brain regions associated with verbal processing.

As we described above, existing computational models have focused exclusively on amodal propositional accounts of problem-solving on the RPM. We propose a complementary computational model that, like Hunt's proposed Gestalt algorithm (1974), uses modal iconic representations of problem inputs. In particular, the affine model that we describe uses pixel-based representations of problem inputs and reasons over these representations using affine transformations and set operations. While this paper focuses on testing the affine model against the Standard Progressive Matrices version of the Raven's test, our more recent work on the affine model has begun to look at the Advanced Progressive Matrices (Kunda, McGreggor, & Goel, 2012). We conclude the introduction with four remarks about the scope of this work.

First, our aim is not to show that the affine model is "better" or "worse" than previous computational models, but rather to explore to what extent a particular set of iconic representations and mechanisms can succeed on a body of RPM problems, just as previous computational models have explored to what extent particular propositional accounts can be successful. We discuss results from the affine model in comparison with other models in order to evaluate how the representational commitments made by such models affect their performance on various subsets of RPM problems.

Second, the affine model demonstrates only one possible instantiation of the use of modal iconic representations for RPM problem solving. The spectrum of possible iconic representations ranges from the type of low-level, pixel-based representation used by the affine model to more complex representations explicitly containing edges, lines, shapes, topological information, etc. One question for further exploration is how models that use other types of iconic representations might perform on the RPM; we have developed one such model, the "fractal model," which solves RPM problems in a manner very different from the affine model while still using iconic, pixel-based representations (McGreggor, Kunda, & Goel, 2010).

Third, while the affine model does not seek to provide an account of or model all of the microstructures and microprocesses of human visual cortical processing, the operations used by the affine model (affine transformations and set operations) are mathematically grounded for general forms of imagery or visualization and are based upon evidence from studies of mental imagery. Both affine transformations and set operations can be formally defined as general types of transformations over any two-dimensional plane figures, whether pixels, edges, shapes, or otherwise. These types of operations, or operations that are computationally isomorphic, have been found to play a role in mental imagery tasks ranging from mental rotation (Shepard & Metzler, 1971) and scanning (Kosslyn, Ball, & Reiser, 1978) to image addition and subtraction (Brandimonte, Hitch, & Bishop, 1992).

Fourth, while the affine model was designed to use forms of inference similar to those evidenced by studies of mental imagery in humans, not all elements of the model are intended to be interpretations of human cognitive processing on the RPM. The primary intent of the model is to evaluate whether the content of the proposed knowledge representation is sufficient for solving RPM problems, using forms of inference that are cognitively plausible, even though certain aspects of the overall process may not be. Thus, the affine model represents a *content* model rather than a *process* model of how humans might solve RPM problems using iconic visual representations.

This work builds upon a long line of research on analogical reasoning. In earlier work, we showed how functional and causal knowledge of physical systems enables analogical reminding and transfer in both within-domain analogies (Goel, Bhatta, & Stroulia, 1997; Goel & Chandrasekaran, 1988) and cross-domain analogies (Goel & Bhatta, 2004; Griffith, Nersessian, & Goel, 2000). In that work, functional and causal knowledge was represented propositionally.

In later work, we showed that visual knowledge and reasoning alone can address some classes of analogy problems that had been assumed to require causal knowledge and reasoning (Davies & Goel, 2001; Davies, Goel, & Yaner, 2008). We also showed how visual analogies can account

for several aspects of creative problem solving in scientific discovery (Davies, Nersessian, & Goel, 2005) and engineering design (Davies, Goel, & Nersessian, 2009). However, this work used propositional representations; while the content of knowledge was visuospatial, the form of representation was still propositional.

2. Affine model for problem-solving on the RPM

The affine model uses representations consisting of two-dimensional arrays of grayscale pixels, with each pixel associated with a single intensity value. These pixel-based representations are iconic in that they preserve a spatial correspondence with the patterns of light and dark areas on the actual test problem inputs. They are modal in that they remain in the same pixel-based format that was generated when test problems were scanned using a digital scanner.

Specifically, the inputs to the affine model for a given RPM problem are sets of images that represent the individual matrix entries and answer choices as presented in the original RPM test booklet. For the 2×2 problem in Fig. 1, the inputs to the affine model are the images shown in Fig. 3, where $\mathbf{m_{ij}}$ refers to the entry at row **i** and column **j** of the matrix, and $\mathbf{a_1}$ through $\mathbf{a_n}$ represent the **n** answer choices given at the bottom of the problem. The output of the affine model is a single number between **1** and **n**, denoting its chosen answer.

2.1. High-level approach

At a high level, the basic approach used by the affine model is to:

(1) Inspect the matrix portion of an RPM problem to determine what relationship is present among the existing matrix entries.
(2) Using this relationship, generate a predicted answer in the form of an image for what entry might occur in the empty spot in the matrix.
(3) Compare the predicted answer to each given answer choice and select the choice that is most similar to the prediction.

The relationship that the affine model attempts to determine in Step 1 is an image transformation that best

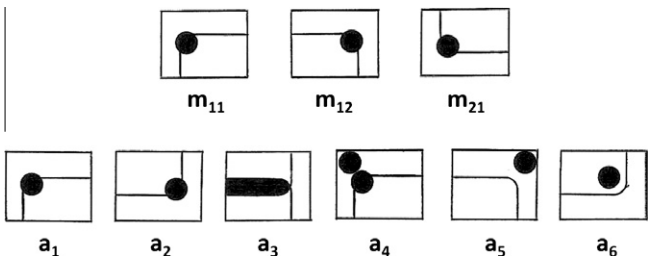


Fig. 3. Imagistic representation of the RPM problem shown in Fig. 1, fed as input into the affine model.

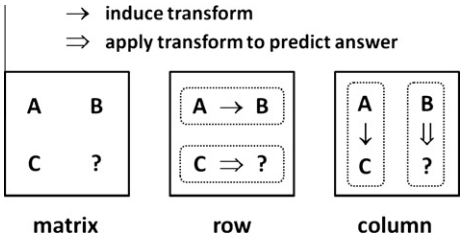


Fig. 4. Schematic illustration of transformations considered by the affine model for a 2×2 RPM matrix.

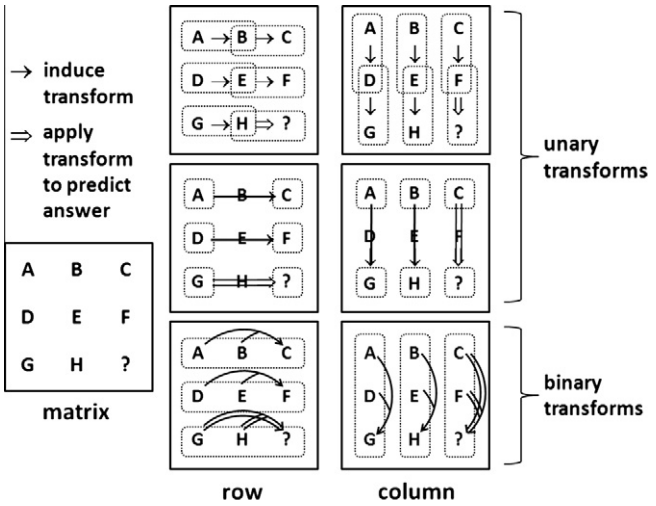


Fig. 5. Schematic illustration of transformations considered by the affine model for a 3×3 RPM matrix.

accounts for the variation among entries in any individual row or column of the matrix. In its present implementation, the model examines both rows and columns before selecting the single best-fit row or column in the matrix. In Step 2, the model applies this same transformation to whichever incomplete row or column is parallel to the first, in order to generate its predicted answer. In these two steps, the affine model is making two implicit assumptions about the structure of RPM problems: (1) entries in a single row or column of the matrix are related according to some image transformation, and (2) parallel rows or columns are analogous in that they share the same image transformation.

Schematic illustrations of which entries the affine model uses in Step 1 to induce row or column transformations are given in Figs. 4 and 5 for 2×2 and 3×3 matrices, respectively. These illustrations show which parallel incomplete rows or columns are used together with the induced transformation to generate the predicted answer in Step 2. For example, looking at a 2×2 matrix as shown in Fig. 4, the model might induce a row transformation relating entries **A** and **B** and then apply this transformation to element **C** to predict an answer. Alternately, the model could try to induce a column transformation in the same manner, first relating entries **A** and **C** and then applying the induced transformation to entry **B**.

For 3×3 matrices, the set of possible transformations is much larger, as there are eight matrix entries to consider instead of just three. Beyond considering unary transformations as in the 2×2 case, i.e. transformations converting a single given image into a single transformed image, 3×3 matrices present the possibility of binary transformations, i.e. transformations converting two given images into a single transformed image. For a 3×3 matrix, looking at row transformations, the model might induce a unary row transformation between adjacent entries **A** and **B** or adjacent entries **B** and **C**, and then apply this transformation to entry **H** to predict an answer, as shown in the top "row" matrix in Fig. 5. Or, the model might induce a binary row transformation relating all three entries **A**, **B**, and **C**, and then apply this transformation to entries **G** and **H**, as shown in the bottom "row" matrix in Fig. 5. As with 2×2 matrices, all of these transformations for 3×3 matrices can be induced either across rows or along columns.

The illustrations of transformations shown in Figs. 4 and 5 represent not all possible relationships among entries in the matrix but merely one subset of such relationships that the current implementation of the affine model was designed to consider. These particular relationships were chosen to encapsulate a problem-solving strategy that focuses on single within-row or within-column relationships and assumes that parallel sets of entries share the same transformations.

Therefore, for a given RPM problem, the affine model proceeds by first inducing all possible transformations for the matrix, both row-wise and column-wise, according to the groupings shown in Fig. 4 or Fig. 5. The transformation induction process is described in more detail below. Each induced transformation carries with it a measure of "fitness" that varies between 0.0 and 1.0 to indicate how well that particular transformation fits its associated row or column, where 0.0 indicates a poor fit and 1.0 indicates a perfect fit. The affine model selects that transformation and associated row/column that has the highest measure of fitness, which completes Step 1.

In Step 2, the model applies this transformation to the appropriate incomplete row/column to predict an answer. Finally, in Step 3, the predicted answer is compared in turn to each given answer choice according to a similarity measure, which is also described below. The choice yielding the highest similarity value is chosen as the model's answer.

2.2. Best-fit image transformations

We now describe the induction process for unary transformations (e.g. converting image **A** to image **B**), with the detailed algorithm given in Table 1.

To begin, suppose we have two images **A** and **B**. We wish to induce a transformation that represents the change in going from **A** to **B**. This process is akin to image registration, in which two images are aligned according to some criteria that ultimately enable a "best-fit" correspondence to be found between the two images. In image registration, a correspondence between two images is found by matching features between the images, and any remaining differences are modeled as a combination of various types of geometric deformations and/or color transformations (Zitová & Flusser, 2003). While image registration is typically performed on real-world images, we adapt this approach for the affine model's transformation induction process, as it seems well able to capture differences between black-and-white line drawings of the type found in RPM problems.

Table 1
Algorithm for calculating best-fit composite transformation between image pair.

For each base affine transform $\mathbf{t_i}$:
I. Apply $\mathbf{t_i}$ to image **A** to create image $\mathbf{t_i(A)}$
II. Search all possible translation offsets between images $\mathbf{t_i(A)}$ and **B** to find single offset (**x**, **y**) yielding highest similarity between them
III. Determine image composition operation $\oplus$ and operand **X** as follows:
- Calculate similarity **s** between image $\mathbf{t_i(A)_{(x,\ y)}}$ and image **B**
- Determine image composition operation and operand as follows:
  - If $\mathbf{\Sigma(A\text{-}B) = 0}$, then $\oplus$ and **X** are null
  - If $\mathbf{\Sigma(A\text{-}B) \leqslant \Sigma(B\text{-}A)}$, then $\oplus$ refers to image addition and $\mathbf{X = B - t_i(A)_{(x,\ y)}}$
  - If $\mathbf{\Sigma(A\text{-}B) > \Sigma(B\text{-}A)}$, then $\oplus$ refers to image subtraction and $\mathbf{X = t_i(A)_{(x,\ y)} - B}$.

The composite transformation $\mathbf{T_i}$ is thus defined as precisely the transformation that changes image **A** into image **B**:

$$\mathbf{T_i(A) = t_i(A)_{(x,y)} \oplus X = B}$$

The similarity value **s** represents how well this transformation fits images **A** and **B**

$\mathbf{T_i}$ applied to a general image **Z** can then be specified as:

$$\mathbf{T_i(Z) = t_i(Z)_{(x,y)} \oplus X}$$

In particular, the affine model defines a composite transformation between two images as a combination of two geometric transforms and one color-based transform:

(1) A base affine transform **t** (e.g. rotation, reflection, etc.).
(2) A translation (**x, y**).
(3) A pixel-wise composition operation $\oplus$ (e.g. addition, subtraction) together with a composition operand **X**, which consists of another image

The affine model contains a finite set of base transforms which, for simplicity, are restricted to rectilinear rotations and reflections. Affine transformations such as shearing and scaling are not included, nor are other types of geometric image deformations.

To induce a composite transformation between two images, the affine model first uses a template-matching scheme to search across all possible base transforms and

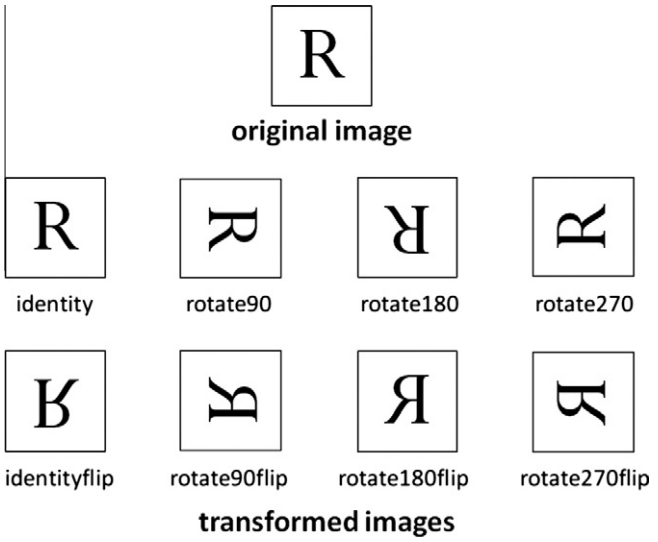


Fig. 6. Eight base unary affine transforms used by the affine model for 2×2 and 3×3 matrices.

translations to find the combination of these two geometric transforms that results in the best correspondence between image **A** and image **B**. Then, given these particular geometric transforms, any remaining image discrepancies are accounted for by defining pixel differences between the two images as comprising the operand of an image composition operation, namely pixel-wise addition or subtraction. Which type of operation is selected depends on whether there are a greater number of pixels being added to or subtracted from image **A** to arrive at image **B**.

Then, the combination of these three transforms—base transform, translation, and image composition—is defined to be the best-fit composite transformation between image **A** and image **B**. The degree of "fit" (i.e. the strength of the discovered correspondence) is defined as the similarity value found during the template-matching process.

Inducing binary transformations (e.g. converting images $\mathbf{A_1}$ and $\mathbf{A_2}$ to image **B**) is a straightforward extension of this process. In particular, the algorithm for inducing a binary transform is identical to that shown in Table 1, except "image **A**" is first created by combining images A_1 and A_2 using a candidate binary transform.

2.3. Base transforms

The base unary transforms (i.e. transforming image **A** into image **B**) used by the affine model during the induction of composite transformations are drawn from the set of image operations that fall under the category of affine transformations (hence the name of the model), and in particular are restricted to orthonormal transformations only (i.e. rotation and reflection, combined with translation). In addition to the fact that affine transformations are a well-defined and thoroughly-studied type of image operation, there is evidence that human visual processing can apply affine transformations like scanning (i.e. translation), zooming (i.e. scaling), and rotation to mental images, or at least operations that are computationally isomorphic (Kosslyn, Thompson, & Ganis, 2006; Shepard & Metzler, 1971).

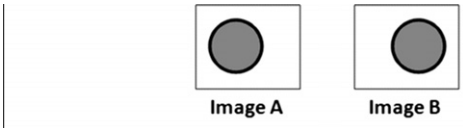
The affine model presently uses the eight base unary transforms shown in Fig. 6, which comprise all possible rectilinear rotations and reflections. In addition, for 3×3 matrices, as mentioned earlier, the larger number of matrix entries introduces the possibility of using binary image transforms instead of unary ones (i.e. transforming images $\mathbf{A_1}$ and $\mathbf{A_2}$ into image **B**). The base binary image transforms used by the affine model are drawn from set composition operations to capture notions of image union, intersection, subtraction, etc., and are implemented at the pixel level as maximums, minimums, and differences of grayscale intensity values. Fig. 7 illustrates the five base binary transforms used by the model.

2.4. Visual similarity

The same similarity measure is used by the affine model in Step 1, for template matching during the transformation induction process, and also in Step 3, to select the final answer choice based on the predicted answer image. This measure is adapted from Tversky's (1977) ratio model of similarity:

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

In this equation, f represents some function over features in each of the specified sets A and B. The constants α and β



Transformation	Notation	Pixel operation on pixels $p_a \in A$ and $p_b \in B$	Resulting image
union	A ∪ B	$\max(p_a, p_b)$	
intersection	A ∩ B	$\min(p_a, p_b)$	
subtraction	A − B	$p_a - p_b$	
back-subtraction	B − A	$p_b - p_a$	
exclusive-or	A xor B	$\max(p_a, p_b) - \min(p_a, p_b)$	

Fig. 7. Five base binary set transforms used by the affine model for 3×3 matrices.

are used as weights for the non-intersecting portions of A and B. If α and β are both set to 1.0, this equation becomes:

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

For calculating the similarity measure, each feature is defined to be a single pixel, and intersection, union, and subtraction operations are defined as the maximums, minimums, and differences, respectively, of the pixels' gray-scale intensity values. The functions f over sets of pixels are defined as simple summations of the feature comparison values over the entire image.

The particular formulation of Tversky's ratio model used by the affine model makes one important assumption about pixels, which is that they can be treated as independent features within the pixel sets represented by images **A** and **B**. While this notion of pixel independence is a strong simplification, it matches the assumptions made by basic template theories of visual similarity that define similarity based purely on evaluations of the extent of overlapping figural units (Palmer, 1978), which in our case are individual pixels.

2.5. A detailed example

We present one detailed example of the operation of the affine model, using the sample RPM problem shown in Fig. 1. First, the original problem image is broken into the constituent matrix entry and answer images, as shown in Fig. 3. Then, as shown in Fig. 4, there are two possible combinations of elements that are used to induce transformations: the elements across the first row and the elements down the first column. The base transforms used in the induction process are the eight rotations/reflections shown in Fig. 6.

For the top row and for the first column, the best-fit composite transformation $\mathbf{T_i}$ is calculated by the model according to the algorithm shown in Table 1. The resulting similarity values from these calculations are given in Table 2. Once these similarity values have been calculated, the transformation yielding the highest similarity is chosen as the defining transformation for the matrix. In this case, it is the rotate180-flip transform as applied to the images in the first row of the matrix, which yields a similarity value of

Table 2
Matrix similarity calculations for the example problem shown in Fig. 1.

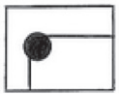
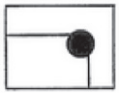
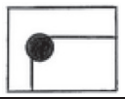
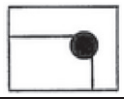
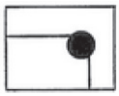
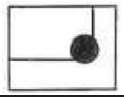
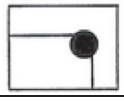
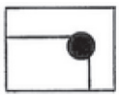
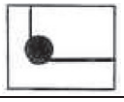
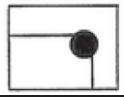
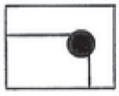
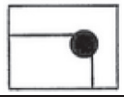
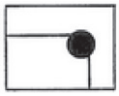
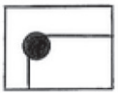
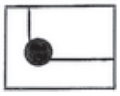
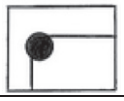
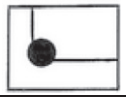
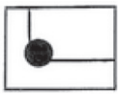
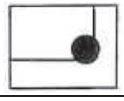
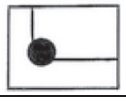
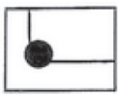
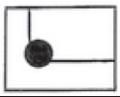
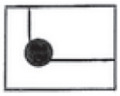
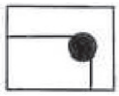
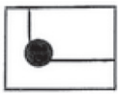
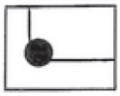
original images	base transform	first image transformed	second image	s	Σ(A-B)	Σ(B-A)
first row:  to 	identity			0.334	226.7	218.5
	rotate90			0.292	250.2	247.8
	rotate180			0.536	120.4	122.4
	rotate270			0.262	269.6	267.0
	identity-flip			0.318	235.9	229.4
	rotate90-flip			0.253	274.2	270.7
	rotate180-flip			0.697	59.5	58.7
	rotate270-flip			0.259	271.2	268.5

Table 2 (*continued*)

first column:  to 	identity			0.438	173.7	158.4
	rotate90			0.255	275.0	263.0
	rotate180			0.323	236.6	213.3
	rotate270			0.311	242.2	228.7
	identity-flip			0.608	104.6	86.3
	rotate90-flip			0.261	272.1	256.8
	rotate180-flip			0.289	254.9	234.1
	rotate270-flip			0.256	274.7	261.8

0.697. Then, for this particular transformation, the image composition operand is determined to be subtraction, as there are more pixels that are in **A** but not in **B** than vice versa, i.e. $\Sigma(\mathbf{A}-\mathbf{B}) > \Sigma(\mathbf{B}-\mathbf{A})$. In other words, the second image **B** roughly equals the first image **A** transformed and minus some pixels.

The predicted answer image is generated by taking the first image from the second row, applying the rotate180-flip transform, and subtracting the same pixels that represent the difference between the images in the top row. In this particular case, the first row images are fairly closely matched, and so the pixels that are subtracted are few in number but not zero, due to slight imperfections in the input images. Finally, the predicted answer image is compared to each of the answer choices, as shown in Table 3. The most similar answer choice is selected as the affine model's final answer, which is answer number 2, with a similarity value of 0.503.

2.6. Model configurations

We implemented three different configurations of the basic affine model described above.

Standard configuration: The standard configuration of the affine model used the Tversky ratio model of similarity, given in Eq. (2), and solved problems based on the single best-fit transform across any matrix row or column, as described in Table 1.

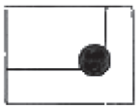
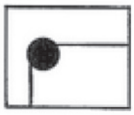
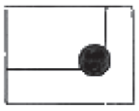
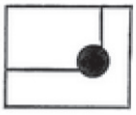
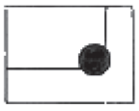
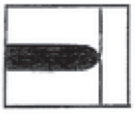
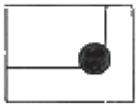
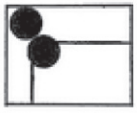
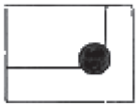
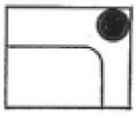
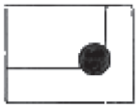
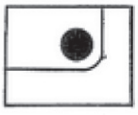
SSD configuration: In order to investigate the effect of using different formulations of visual similarity, this configuration uses a sum-squared-differences (SSD) measure of similarity instead of the Tversky similarity measure, defined as:

$$\text{similarity}(A,B) = \frac{1}{1+\sum(A-B)^2} \tag{3}$$

Note that the model takes the reciprocal of one plus the sum of squared differences between pixel intensities in order to convert the usual SSD measure of difference, varying from 0 to positive infinity, into one of similarity, varying from 0.0 to 1.0.

Aggregate configuration: The standard affine configuration determines the best-fit image transformation for a matrix by searching among all possible base transforms and among all sets of entries listed in Figs. 4 and 5. For 3×3 problems, an alternate strategy was implemented that selects the best-fit image transformation by searching among the possible base transforms with fitness values averaged for sets of entries across all complete rows or columns. After the best-fit base transform has been chosen

Table 3
Answer similarity calculations for the example problem shown in Fig. 1.

predicted answer image	answer choice images	S
		0.257
		0.503
		0.256
		0.211
		0.265
		0.277

based on this aggregate fitness value, the single best-fit row or column is used together with the corresponding partial row/column, as in the standard algorithm, to generate an answer prediction. This configuration uses the standard Tversky similarity measure.

3. Results

The Standard Progressive Matrices (SPM) consists of 60 problems divided into five sets of 12 problems each, labeled Sets A–E, with problems increasing in difficulty both within and across sets. We tested the affine model on all 60 problems from the SPM. We began with a paper copy of the test and scanned it to create digital images of each problem. Each of the 60 problem images was manually rotated so that the matrix lay squarely along horizontal and vertical axes. Then, each problem image was subdivided into individual input images, as shown in Fig. 3.

Each 2×2 problem (Sets A and B) had three matrix image inputs and six answer choice images, with each image being roughly 135 by 90 pixels in size. Each 3×3 problem (Sets C–E) had eight matrix image inputs and eight answer choice images, with each image being roughly 80 by 60 pixels in size. Images were represented as arrays of grayscale intensity values, where a value of 0.0 corresponded to white and a value of 1.0 corresponded to black. To reduce the effects of image noise, the affine model converted any pixel intensity value less than 0.5 to a value of 0.0 throughout its computations.

The standard configuration of the affine model correctly solves 35 of the 60 problems on the SPM. For typically developing children in the US, this total score corresponds to the 75th percentile for 8-year-olds, the 50th percentile for 10-year-olds, and the 25th percentile for 12½-year-olds (Raven et al., 2003).

The SSD configuration of the affine model correctly solves 33 problems. It misses 6 problems that the standard configuration solves correctly—one from Set A, four from Set C, and one from Set E—and solves 4 problems that the standard configuration misses—all from Set E. A total score of 33 corresponds to the 75th percentile for 8-year-olds, the 50th percentile for 9-year-olds, and the 25th percentile for 11½-year-olds.

The aggregate configuration of the affine model correctly solves 38 problems. It misses two problems that the standard configuration solves correctly—one from Set C and one from Set E—and solves five problems that the standard configuration misses—two from Set C and three from Set E. A total score of 38 corresponds to the 75th

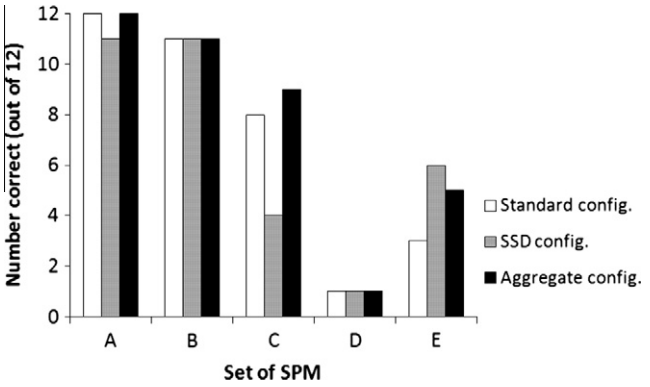


Fig. 8. Scores achieved by three different configurations of the affine model on the SPM, broken down by Sets A–E.

percentile for 9-year-olds, the 50th percentile for 11-year-olds, and the 25th percentile for 14-year-olds.

Breakdowns of these total scores across sets are shown in Fig. 8. As this figure shows, all three configurations of the affine model perform fairly well on Sets A and B while performing poorly on Set D. On Set C, the SSD configuration appears to be at a disadvantage, while on Set E, the standard configuration fares more poorly than the other two configurations. We discuss some possible reasons for these patterns of performance in the following section.

4. Discussion

4.1. The affine model: Results and analysis

The affine model performs surprisingly well on the SPM, given its limited repertoire of image operations. The standard configuration of the model, which uses only eight basic affine transformations along with five basic image composition transformations, correctly solves over half of the problems on the SPM, which is well above a random-guessing baseline.[3]

One noticeable feature of the results is that all three configurations of the affine model do poorly on Set D, solving only one out of twelve problems. This low performance is even more striking when compared to expected score distributions for human test-takers: according to published norms, most people who receive total scores in the mid-30s have correctly solved 6 or 7 problems in Set D (Raven et al., 2003). Furthermore, Set E is typically more difficult for people than Set D, but all three configurations of the affine model do better on Set E than they do on Set D.

Upon inspection, it appears that nearly all of the problems in Set D require two types of manipulations of matrix elements: (1) permutations of entries across rows or columns and/or (2) segmentation of a single matrix entry into multiple elements that follow different transformation rules. For example, a problem might have three inner shapes that are permuted across rows and columns and three outer shapes that remain constant across rows, as illustrated in Fig. 9. The affine model cannot currently account for these types of transformations, though there is no *a priori* reason why such transformations could not be implemented using iconic representations. Permutations, which are implemented in propositional systems like the Carpenter et al. (1990) model as "distribution" rules over elements, could be handled by the affine model by translating entire rows or columns of the matrix to align identical entries or by considering diagonal sets of matrix entries in addition to rows and columns when inducing transformations. Segmentation or partitioning of images into subsets of objects or features could be performed by iteratively seeking transformations to successively explain differences between various subsets of pixels in each matrix entry, until no pixels remain to be explained.

Both of these types of operations would likely be necessary for a computational model to do well on Set D, as well as on the more difficult problems in the Advanced Progressive Matrices test. The Carpenter et al. (1990) model did not perform automatic image segmentation, as inputs were hand-coded (and thus hand-segmented) into propositional features. The Cirillo and Ström (2010) model also did not perform automatic image segmentation; inputs were redrawn by hand as segmented vector graphics before being passed into the model, and then an automated system extracted propositions from these representations. Lovett et al. (2010) also recreated test problems as segmented vector graphics, but their system did have the ability to re-group and re-segment discrete shapes and edges within the vector graphics representations. A question for future

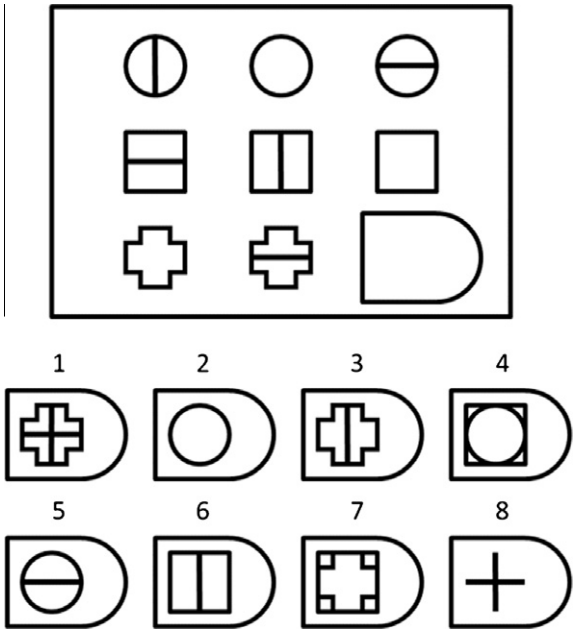


Fig. 9. Example problem showing permutations as well as multiplicity of elements in individual matrix entries.

[3] Simple probability calculations suggest that a random guessing strategy on the SPM would yield an average expected score of 8.5 out of 60 problems.

work is how automated image processing techniques might be applied to perform image segmentation of RPM problems, and what background knowledge is needed regarding the identities of shapes and other visual entities in order to perform such segmentation.

Another feature of the results is the flip-flop in performance between the standard and SSD configurations of the affine model on Sets C and E. On Set C, the standard configuration solves eight problems correctly, while the SSD configuration solves only four. On Set E, in contrast, the SSD configuration solves six problems, while the standard configuration solves only three. How does changing the similarity measure affect performance on these sets in such different ways? A closer look at each similarity measure suggests one possible answer.

Consider the image pairs shown in Fig. 10. For pairs **AB** and **CD**, the number of pixels that are different between the images within each pair is the same—two pixels—but the amount of common pixel content that is shared is different—four pixels in pair **AB** and only two pixels in pair **CD**. The Tversky measure, given in Eq. (2), privileges matches that share more pixel content, and so images **AB** yield a higher similarity value than do images **CD**. In contrast, the SSD similarity measure, given in Eq. (3), effectively ignores any pixel content that is shared; similarity is calculated only as a function of pixels that are different. Thus, the SSD measure yields identical similarity values for image pairs **AB** and **CD**, because within each image pair, there are two mismatched pixels. The opposite pattern can also occur: for image pairs **EF** and **GH**, the Tversky measure yields identical similarity values, but the SSD measure prefers pair **EF**, because **EF** has only two mismatched pixels, whereas **GH** has three mismatched pixels.

Looking at the problems in Sets C and E on the SPM, most of the problems in Set C involve shapes that are filled solidly or with textures. The Tversky similarity measure appears to be more successful at aligning these types of shapes than is the SSD measure, as image matches that share large swaths of pixel content will receive higher similarity ratings. The problems in Set E, in contrast, are mostly composed of thin edges and lines, for which the SSD similarity measure seems better able to capture fine differences between edge alignments in a pair of images.

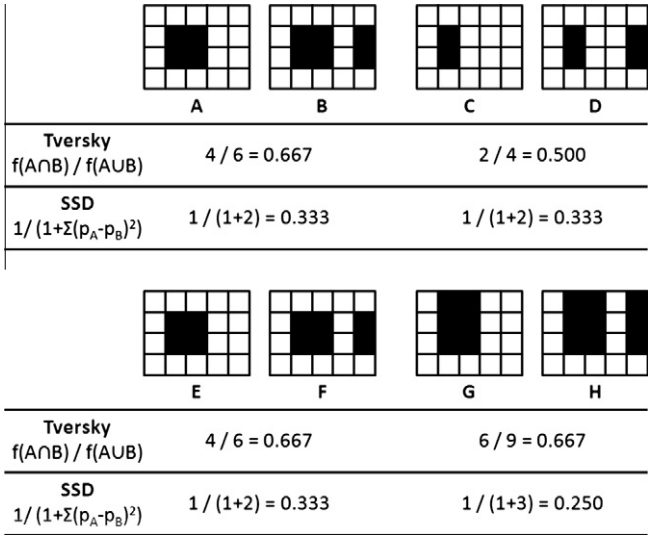


Fig. 10. Illustration of differences between Tversky (1977) and SSD similarity measures, as applied to pixel-based images.

Calculating similarity is a central facet of the affine model, and would be of any model using modal iconic representations. Our experiments with the affine model using two different formulations of visual similarity show that, while there can be non-trivial effects stemming from biases inherent in various similarity measures, the affine model as a general approach can accommodate different similarity measures without significant change in its problem-solving power. Despite the differences in these similarity measures, both model configurations correctly solve 29 of the same SPM problems.

Across all sets, the aggregate configuration (which also uses the Tversky similarity measure) performs as well or better than the standard configuration. One way to conceptualize why this is so is to consider that SPM problems, especially for 3×3 matrices, contain much redundant information that can be used to find the correct answer. For example, the same transformation often applies from the first entry to the second in a row, from the second entry to the third in that same row, and across the same pairs of entries in each additional row of the matrix. The aggregate configuration takes into consideration this redundancy of information, whereas the standard configuration considers only a single image pair (or triplet) at a time.

4.2. Comparison with other RPM models

Table 4 gives a comparison of the affine model with other computational models of the RPM. The first column identifies the model. The second column indicates whether the representations used are modal iconic representations or amodal propositional representations. The third column shows the type of input received by each model. The fourth column specifies high-level strategy with respect to whether each model (1) predicts the answer according to the information contained in the matrix and then compares this prediction to each answer choice, or (2) guesses each answer choice in turn and evaluates how well it fits into the matrix. Finally, the fifth column gives the scores obtained by each model on various sets of the SPM.

We focus on differences between the two models that use modal iconic representations—the affine model described in this paper, and the fractal model described in McGregor et al. (2010)—and those that use amodal propositional representations—the Carpenter et al. (1990) production system model, the Cirillo and Ström (2010) pattern-matching model, and the Lovett et al. (2010) structure-mapping model.

The two propositional models that have been tested on the SPM, Lovett et al. (2010) and Cirillo and Ström

Table 4
Comparison of five computational models of RPM problem solving.

Model	Type of representation	Inputs to model	High-level strategy	Set of SPM				
				A	B	C	D	E
Carpenter et al. (1990)	Propositional	Hand-coded propositions	Predict answer	n/a; tested against APM only				
Cirillo and Ström (2010)	Propositional	Redrawn vector graphics	Predict answer	n/a	n/a	8	10	10
Lovett et al. (2010)	Propositional	Redrawn vector graphics	Guess-and-check	n/a	44 total (breakdown across sets not given)			
McGreggor et al. (2010)	Iconic	Scanned images	Guess-and-check	11	7	5	7	2
Affine model								
1. Standard				12	11	8	1	3
2. SSD	Iconic	Scanned images	Predict answer	11	11	4	1	6
3. Aggregate				12	11	9	1	5

(2010), both do quite well on Sets D and E when compared against the performance of the iconic models. Part of the reason for this difference may have to do with the image segmentation issue mentioned earlier. Sets D and E both contain complex problems that often involve multiple elements within each matrix entry changing in different ways across the matrix. Neither the affine model nor the fractal model performs any explicit image segmentation; all pixels in each matrix entry are treated as equal, and a single type of image operation is assumed to apply to every pixel. Both of the propositional SPM models receive inputs that have already been segmented into discrete shapes via the redrawing of test problems as collections of vector graphics. Identifying correspondences among these segmented shapes in various rows or columns is certainly a non-trivial reasoning task, and each of these models, as well as the Carpenter et al. (1990) APM model, expend considerable computational effort in discovering shape correspondences. It remains to be seen whether adding image segmentation capabilities to the affine model might boost its performance on Sets D and E up to the level of these propositional models without the need for explicit identification of discrete shapes.

One other interesting aspect of the SPM results is that (as far as we can tell from published findings) only the models using iconic representations have ever attempted Set A of the SPM, which, according to human normative data, is purportedly the easiest set on the test. The problems on Set A of the SPM (see Fig. 11 for examples), are qualitatively different from the problems on Sets B through E in that they resemble pattern-completion problems more than geometric analogy problems. It may be that part of the reason that no propositional models have been tested against Set A is because these types of problems are very difficult to represent using propositions, especially within propositional schemes that focus on representing discrete shapes and attributes. (The APM contains four problems resembling the problems in Set A of the SPM; of these four, only one was attempted by the Carpenter et al. (1990) model, and this one happened to contain discrete elements not unlike those in the geometric analogy type of problem, except with continuous lines added around the elements. No reason was given for the omission of the other three pattern completion problems, though again, according to human normative data, they were supposed to be some of the easiest problems on the test.)

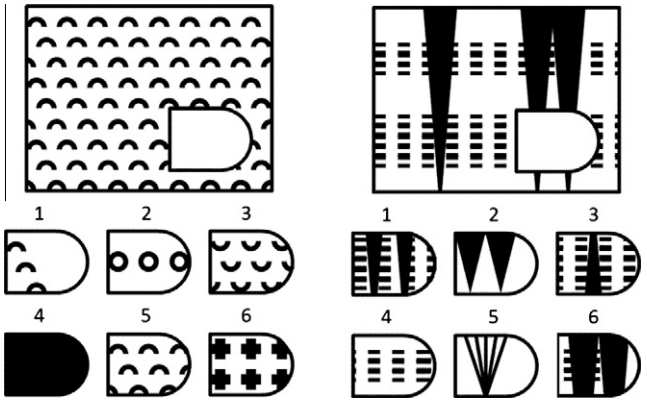


Fig. 11. Examples of the "pattern-completion" type of problems found in Set A of the SPM.

Some of these types of problems likely could be represented propositionally as textures, but such an approach might prove difficult for problems such as that shown on the right of Fig. 11, in which no quadrant of the matrix contains a uniform texture. Furthermore, extracting propositional descriptions of texture directly from an image is in itself a difficult computational task. These problems might also be represented propositionally using a richer vocabulary that includes lower-level elements such as edges and lines (for example, as obtained by the edge segmentation process in the Lovett et al. (2010) model), but this approach could greatly increase the computational complexity of the problem; instead of problems containing two or three or even 10 elements per matrix entry, a single problem like those shown in Fig. 11 might have dozens or even hundreds of elements.

Such problems are very easy to represent using modal iconic representations of the type used by the affine and fractal models; the representation simply consists of the scanned images from the test. In fact, in terms of representation, none of the problems on the SPM are particularly harder or easier to represent than any others using a pixel-based representation. This type of representation seems to be a highly effective choice, as both the affine and fractal models do exceedingly well on Set A of the SPM.

Human factor-analytic studies of the SPM have typically classified these pattern completion problems as

loading on a "gestalt" cognitive factor, in contrast to visuospatial or verbal factors. These data seem to suggest that pattern completion problems may be solved by humans using qualitatively different strategies than those used on the geometric analogy type of RPM problem. The affine model currently solves the problems in Set A using the same mechanisms used on later problems. In particular, the affine model looks at discrete transformations within the problem matrix, i.e. going from one image to another, which is akin to using a rule-based, albeit visual, approach (where the rules are conceptualized as image operations of affine and set transformations). A gestalt approach might differ by looking at the entire problem matrix as a whole, using principles of visual coherence such as symmetry and continuity.

Finally, as listed in the fourth column of Table 4, the high-level strategies chosen by various computational RPM models are not strictly constrained by their choice of representation. The affine model currently uses an answer prediction approach, which is similar to the high-level approach used by the Carpenter et al. (1990) model and by the Cirillo and Ström (2010) model. The fractal model uses a guess-and-check approach, which is similar to the high-level approach used by the Lovett et al. (2010) model. The application of both of these types of strategies in an iterative fashion was proposed by Hunt in both of his RPM algorithms (1974), and studies of human behavior, primarily through eye-tracking analyses, have suggested that humans too use various combinations of prediction and testing in order to arrive at a final answer (Bethell-Fox, Lohman, & Snow, 1984; Hayes, Petrov, & Sederberg, 2011; Vigneau, Caissie, & Bors, 2006). Future work on the affine model will include incorporating a guess-and-check strategy as an alternative or complementary approach to the current answer prediction strategy.

4.3. A note on inputs

Throughout this paper, we have focused our discussion of representations primarily on what type of representation a particular model uses to reason through a given RPM problem. Here, we briefly discuss the types of representations that a model might receive as inputs.

The affine model takes as inputs scanned images from the actual SPM test booklet. Some preprocessing is done on these images; they are manually rotated to correct for rotational misalignments during the scanning process, they are sliced into constituent images for each matrix entry and answer choice, and the images are posterized to remove any light grey pixels, as they are assumed to be noise. Even after these preprocessing steps, however, the images fed as inputs into the affine model are still very noisy; they contain numerous pixel-level artifacts and misalignments from the scanning process, and in addition, the figures in the SPM test booklet are not (at a fine level of detail) as precise as they might appear to the human eye. For example, part of one matrix element that appears to be symmetric about its horizontal axis can be measured and found to be 15% longer on one side than on the other (for example, see answer choice #4 in problem D10 of the SPM). Elements that are clearly meant to appear identical across multiple matrix entries are not exact duplicates of one another; this becomes especially apparent in figures that incorporate textures such as stripes or polka dots.

For these reasons, the similarity values calculated by the affine model are often much lower than one might expect. In the example problem discussed earlier, even though the predicted answer looks very like one of the given answer choices, as shown in Table 3, the calculated similarity between the two images is only 0.503.[4] For the 35 problems correctly solved on the SPM, the average final similarity value calculated by the affine model is 0.599.

One might ask, why not just create "clean" computerized input images to eliminate the imprecision found in scans of the SPM test booklet? We have three reasons for choosing to work with the original scanned images. The first reason is a simple one: given that humans use paper copies of the test, we feel that our models should try to tackle inputs that are as close as possible to the originals. Humans do not receive the benefit of having "cleaned up" versions of RPM problems, and so neither should a computer model.

A second reason has to do with model robustness when faced with low-level representational irregularities. Part of the power of amodal propositional representations comes from their ability to abstract away from the raw pixel level, and, for example, call two squares "identical" despite slight mismatches in size or alignment. We aim to show that methods using modal iconic representations can also achieve similar levels of robustness using calculations of visual similarity at the pixel level, whether or not the inputs have been "cleaned up." The field of image processing regularly deals with noisy, imperfect images, and we wish to maintain some of that realism and take the actual RPM test problems as they come.

The third and most important reason for choosing not to redraw RPM problems is that we feel there is a strong methodological argument against it (whether they are redrawn as vector graphics or even just as more precise raster images). As an example, consider redrawing the shapes shown in Fig. 12. At first glance, these images might appear to be identical, and it would be tempting to create the first circle with stripes and copy it in order to create the second. However, closer inspection will reveal that, although the high-level texture might be described in the same way, at a low level, the images are drastically different—the calculated similarity between these two images is a mere 0.253! While the outer circular outlines are alike, the inner "textured" portion of each circle is almost exactly a negative image of the other.

[4] This example problem was hand-drawn using rulers, stencils, and ink, in order to emulate the level of imprecision found in the actual SPM test booklet.

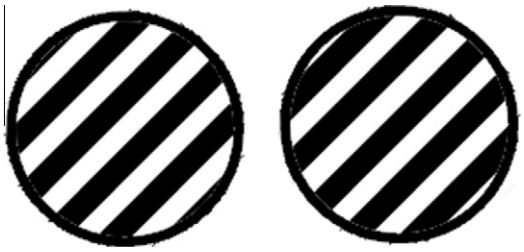


Fig. 12. Illustration of "same" texture with wildly different pixel-level properties.

As another example, specific to redrawing problems as more precise raster images, consider the problem shown in Fig. 13, which was drawn using vector graphics in PowerPoint and then exported as a raster image. Looking just at the top row of matrix elements, and using the set of eight affine base transformations shown in Fig. 6, it becomes apparent that the top-row image transition could equally well be described as a "rotate180flip" transformation (i.e. a reflection about the vertical axis) or as a "rotate270" transformation (i.e. a one-quarter counter-clockwise rotation). It follows that the model *ought* to compute that either of these transformations is equally well-suited, and choose one according to whatever tie-breaker is in place.

However, the actual output of the model depends, in fact, on how the problem was originally created using vector graphics, even after the images have been rasterized. In particular, when recreating this problem using vector graphics in Powerpoint, we took the original, top-left image in the matrix and constructed two different versions of the top-right image. For the first version (the "rotated version"), we took the top-left vector graphic and rotated it 90° to the left. For the second version (the "reflected version"), we took the top-left vector graphic and reflected it about its vertical axis. Then, all of these vector images were rasterized to create input images to feed into the affine model.

Results from calculating similarity values over all base affine transformations for these two versions of the top-row images are shown in Table 5. For the rotated version, the rotate transformation is found to yield the highest image similarity. In contrast, for the reflected version, the rotate180-flip (i.e. reflection) transformation is found to yield the highest similarity. While the slight differences present in the final rasterized images would likely not influence the behavior of a human taking the test, these differences represent enough of a bias that they can completely change the output of a model that uses pixel-based representations, to the point where reconstructing the input using the "correct" transformation leads to a correct answer from the model, while reconstructing the input with a different transformation leads to an incorrect answer.

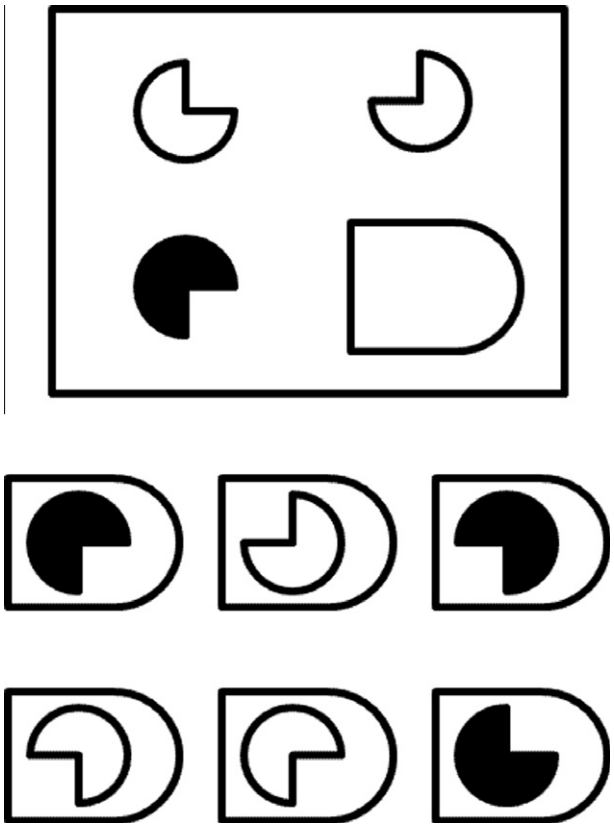


Fig. 13. 2×2 example problem to illustrate impacts of "clean" input images.

As these examples show, when redrawing RPM problems, the specific choices by which "clean" images are created can have a non-trivial impact on the visual

Table 5
Similarity calculations for example problem shown in Fig. 13. Bold values indicate maximum similarity values.

Base transform	Original images	s	Original images	s
Identity	Rotated version:	0.456	Reflected version:	0.439
Rotate90		0.347		0.325
Rotate180		0.449		0.431
Rotate270		**0.884**		0.818
Identity-flip		0.341		0.340
Rotate90-flip		0.458		0.433
Rotate180-flip	to	0.881	to	**0.825**
Rotate270-flip		0.452		0.419

information contained in the problem and thus can significantly alter the output of a computer model. Redrawing could also introduce bias if the drafter has foreknowledge of the computer model to be tested against the problems, as they may consciously or unconsciously redraw problems with the problem-solving algorithm in mind. Lovett et al. (2010) note that for their experiments, one of the SPM test problems was redrawn using a grey line instead of the original dotted line "for simplicity." While humans solving this problem would likely not be much affected by such a change, it does raise questions of when such simplifications are appropriate and when they might, in fact, be materially changing the substance of a problem for a computational system. For all of these reasons, we have deliberately used images scanned directly from the printed SPM test booklet as inputs to our models.

4.4. Conclusion

The question of what sort of mental representations people use to solve particular tasks is central to the study of cognition. Psychology has provided evidence about how tasks are typically solved; however, for many tasks, there may be multiple strategies that *can* be used, and either for particular task variants or for different subsets of the population, different strategies may be at play (Kunda & Goel, 2011).

In the growing cognitive science literature on analogy, several lines of research have explored visual analogies (e.g. Casakin & Goldschmidt, 1999; Clement, 2008; Croft & Thagard, 2002; Davies & Goel, 2008; Evans, 1968; Hofstadter, 1995; Leyton, 2001; Nersessian, 2008; Ojha & Indurkhya, 2009; Stafford, 2001; Yaner & Goel, 2006). Several factors explain this emphasis on visual analogy: for example, the requirements of task and domain, explanations of behavioral data, and consistency with theories of mental imagery. Another important reason is that visual analogies support the construction of representations as well as re-representations, as advocated, for instance, by Indurkhya (1998) and Kokinov (1998). Indeed, Chalmers, French, and Hofstadter (1992) view much of analogy as high-level perception in which representations are constructed rather than assumed as given. The various theories of visual analogy, however, differ along the dimensions of modal/amodal and iconic/propositional representations.

Another good example of the distinction between modal iconic representations and amodal propositional representations comes from a series of findings in the domain of sentence–picture verification tasks (SPVTs). The basic SPVT presents the subject with two stimuli: a sentence or phrase describing some spatial relationship between objects, e.g. "The star is not above the plus," and a picture illustrating the same objects in some arrangement, e.g.  . The subject's task is to make a true/false response depending on whether the sentence and the picture are consistent. Models of how people solve SPVTs have generally relied on the assumption that, given that the two stimuli are encountered in different modalities, they must be represented mentally in some common format in order to make direct comparisons between them (Clark & Chase, 1972). Early models of the SPVT further assumed that this common format was propositional, and these models were able to make predictions about response latencies that seemed well-matched by human behavior (Carpenter & Just, 1975; Clark & Chase, 1972).

However, further studies found evidence that under certain experimental conditions (Tversky, 1975) or for subsets of the population (MacLeod, Hunt, & Mathews, 1978), participants will recode both the sentence and the picture using pictorial encoding rather than propositional encoding. Later studies of the SPVT have confirmed these dual-strategy findings using both behavioral and neuroimaging measures (Coney, 1988; Neubauer & Freudenthaler, 1994; Reichle, Carpenter, & Just, 2000). Clark and Chase (1972) had explicitly discounted the notion of pictorial encoding being used on the SPVT, though their arguments were mainly directed against pictorial encoding being the *only* strategy in use by humans. They had also adopted a limited view of the representational power of pictorial encodings and mental imagery that has since been strongly criticized (Barsalou, 1999).

How does this body of work on SPVTs relate to current research on the RPM? Apart from basic processes of perception and response generation that must take place in any task of this sort, both tasks appear to be amenable to two core reasoning strategies that are qualitatively different in terms of representation, in particular a propositional/verbal representation versus a iconic/pictorial representation. For both the SPVT and the RPM, it seems unlikely that one type of strategy is the "correct" one; rather, they form complementary accounts that are both commonplace in human cognition. In both tasks, strategy variations appear to manifest as between-individual differences, in that certain individuals seem to prefer one strategy over another, as well as within-individual differences, in that single individuals often appear capable of using either strategy. Finally, fMRI studies of both tasks have found evidence for different patterns of brain activation depending on the strategy being employed, and these patterns are consistent with the verbal versus visual nature of each strategy.

A major success from studies of the SPVT has been the ability to connect detailed computational or theoretical models of each type of strategy to specific behavioral predictions about response latencies. While the cognitive science literature on the RPM does propose many detailed models, including the one in this paper, we do not yet have clear and precise behavioral predictions that could be used to identify when an individual might be using a particular strategy. The issue is further complicated by the fact that qualitative differences in RPM strategies seem to lie along several orthogonal dimensions, from differences in representation (e.g. visual versus verbal strategies) to differences in high-level approach (e.g. answer

prediction versus guess-and-check) and potentially many others (Vigneau et al., 2006). The closest example of a model-to-behavior prediction comes from the Carpenter et al. (1990) paper, which makes predictions about the numbers of eye fixations that subjects might make on different types of problems. While the experimental results presented are consistent with the model in certain respects, there is ambiguity in terms of which portions of the model might be validated by such data. For example, while the eye-gaze data might elucidate a participant's high-level approach (i.e. answer prediction versus guess-and-check) or whether they adopt a rows-first or columns-first tactic, no direct relationship exists between the eye gaze data and the use of a purely propositional strategy; the data might be equally well fit by a model that uses iconic representations.

Thus, in order to generate useful predictions about human RPM performance, an RPM model must be precise in its commitments to the particular type of strategy variation it is attempting to discern. We identify five potential areas for RPM predictions—reaction time, eye-tracking, accuracy, error patterns, and neuroimaging—and briefly discuss their relationships to the affine model presented in this paper.

Reaction time: Measures of reaction time were the primary means of validation for the early SPVT studies discussed above. However, the RPM represents a much more complex task than the SPVT; while the SPVT contained discrete and sequential stages that made it easy to predict reaction times using simple additive assumptions, the RPM, as conventionally given, is not readily broken down into stages, as it involves a single presentation of the entire problem and all answer choices at the same time. In addition, people likely do not solve the RPM in a unidirectional and sequential fashion. Hunt's (1974) RPM algorithms contain numerous loops and iterative sub-processes. Furthermore, the affine model in particular does not represent a detailed process model of the problem-solving procedure. It may be that the affine model can be used to predict reaction time through some proxies, such as number of transforms used for a given problem, but this approach might be confounded by whether the predictions are a function of item difficulty, independent of representational modality. Thus, reaction time studies seem a difficult path for testing representation modality-based models of RPM problem-solving.

Eye-tracking: Eye-tracking studies of human RPM performance have the ability to elucidate patterns of visual attention. Eye-tracking data have already been used to study people's high-level approach to solving RPM problems, in terms of whether they predict an answer or guess-and-check among the available answers (Bethell-Fox et al., 1984; Hayes et al., 2011; Vigneau et al., 2006). However, it is not apparent how this type of eye-gaze data could illuminate what type of representation modality an individual is using, as the same attentional strategy might be feasible to use with various representations, and a single representation modality might be amenable to different attentional strategies.

Accuracy: Accuracy data may be useful in distinguishing among various representational strategies, particularly through the identification of problem subtypes, i.e. a particular problem may be very easy, very difficult, or even impossible to solve using a particular type of representation. Identifying or developing such problem subtypes, however, is no simple matter. One approach is to base subtypes on human performance data; however, one potential confound for this approach lies in the difference between how problems are *typically* solved versus how they *can* be solved (Kunda & Goel, 2011). Many studies of the RPM, such as factor analytic studies, identify problem types based on how a majority of their participants appear to solve particular problems; however, these problem classifications are not sufficient to show whether, for instance, a "verbal-analytic" problem might in fact be solvable using visual representations, even if not typically solved that way. Another approach might identify subtypes based on the results of a computational model; however, while a particular model can show that a representation is sufficient for solving a particular problem, it is very difficult to rule out representations as being potentially successful. A different approach that may be more promising is to classify or develop problems based on a structured understanding of how problem components interact with difficulty levels—in other words, constructing problems specifically to tax certain cognitive processes (Primi, 2001; Meo, Roberts, & Marucci, 2007). Primi (2001) gives an excellent example of how two problems can be identical in terms of the numbers and types of rules and elements that they contain, which would make them equally easy to solve using a propositional strategy (after encoding), but vary significantly in terms of their perceptual organization, which would lead one problem to be easily solvable using a visual strategy but the other much more difficult. This approach could begin to distinguish between visual and verbal strategy use in humans, although care would need to be taken to account for perceptual operations that contribute to the encoding portion of verbal strategies.

Error patterns: Different representational strategies may well lead to observable differences in patterns of errors on the RPM, particularly in terms of which wrong answer might be chosen for a problem answered incorrectly. We are currently attempting to discern strategy differences between typically developing individuals and individuals with autism by looking at patterns of errors made on the SPM by human subjects and by the affine model. Part of this work will include refining the affine model to specify how final answer choices are selected based on computed similarity values, yielding probability distributions over the answer choices that can then be compared to human data. Results from this work are still preliminary (Kunda, Soulières, Mottron, & Goel, 2011). Another promising approach for examining error patterns is to artificially construct RPM items such that different strategies clearly bias

test-takers towards one answer choice versus another. Kirby and Lawson (1983) developed a set of items, based on Hunt's (1974) analysis of Gestalt and Analytic strategies, which had ambiguous answer choices of this kind.

Neuroimaging: Much neuroimaging data on human RPM performance has seemed consistent with the idea of a distinction between visual and verbal strategies (Prabhakaran et al., 1997; Soulières et al., 2009). Insofar as fMRI data can distinguish between broad areas of brain activation for visual versus verbal cognitive processing, neuroimaging will continue to be very useful in the study of representational strategy differences on the RPM. However, these data may be too coarse-grained to make finer distinctions about details of various strategies, for instance in distinguishing among different visual strategies or characterizing particular aspects of a mental representation.

In conclusion, ever since Hunt's (1974) work suggesting that the RPM is amenable to solution using different types of representations, there has been a growing body of behavioral and neuroimaging evidence suggesting that human RPM strategies do differ according to whether they use visual or verbal representations. For both practical and theoretical considerations, it is essential that we continue to investigate and discuss a diverse range of RPM strategies in order to better understand the entire collection of problem-solving processes that humans use when solving the test. To this body of work, we add evidence from the affine model showing that visual RPM strategies using modal iconic representations are computationally feasible.

Acknowledgments

This work has benefited from many discussions with Agata Rozga, a developmental psychologist in the School of Interactive Computing at the Georgia Institute of Technology in Atlanta, GA, USA. We are grateful to the US National Science Foundation for its support of this research through NSF (RI) Grant #1116541, titled "Addressing visual analogy problems on the Raven's intelligence test." We also thank the US Office of Naval Research, who supported this work through an NDSEG graduate fellowship, and the NSF GRFP graduate fellowship program. Finally, we thank the editors and reviewers of this journal for prompt and helpful reviews.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660.

Berker, E., & Smith, A. (1988). Diaschisis, site, time and other factors in Raven performances of adults with focal cerebral lesions. *The International Journal of Neuroscience, 38*(3–4), 267–285.

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*, 205–238.

Bölte, S., Dziobek, I., & Poustka, F. (2009). Brief report: The level and nature of autistic intelligence revisited. *Journal of Autism and Developmental Disorders, 39*(4), 678–682.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Manipulation of visual mental images in children and adults. *Journal of Experimental Child Psychology, 53*(3), 300–312.

Bringsjord, S., & Schimanski, B. (2003). What is artificial intelligence? Psychometric AI as an answer. In *IJCAI* (Vol. 18, pp. 887–893).

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82*(1), 45–73.

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), 404–431.

Casakin, H., & Goldschmidt, G. (1999). Expertise and the use of visual analogy: Implications for design education. *Design Studies, 20*, 153–175.

Chalmers, D., French, R., & Hofstadter, D. (1992). High-level perception, representation and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical Artificial Intelligence, 4*, 185–211.

Cirillo, S., & Ström, V. (2010). *An anthropomorphic solver for Raven's progressive matrices* (No. 2010:096). Goteborg, Sweden: Chalmers University of Technology.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Dordrecht: Springer.

Coney, J. (1988). Individual differences and task format in sentence verification. *Current Psychology, 7*(2), 122–135.

Croft, D., & Thagard, P. (2002). Dynamic imagery: A computational model of motion and visual analogy. In L. Magnani & N. Nersessian (Eds.), *Model-based reasoning: Science, technology, and values* (pp. 259–274). New York: Kluwer Academic/Plenum Publishers.

Davies, J., & Goel, A. (2001). Visual analogy in problem solving. In *Proceedings of the 17th international joint conference on, artificial intelligence (IJCAI-01)* (pp. 377–382).

Davies, J., & Goel, A. (2008). Visual re-representation in creative analogies. *Open AI Journal, 2*, 11–20.

Davies, J., Goel, A., & Yaner, P. (2008). Proteus: Visuospatial analogy in problem solving. *Knowledge-Based Systems, 21*(7), 636–654.

Davies, J., Goel, A., & Nersessian, N. (2009). A computational model of visual analogies in design. *Journal of Cognitive Systems Research, 10*, 204–215.

Davies, J., Nersessian, N., & Goel, A. (2005). Visual models in analogical problem solving. *Foundations of Science, 10*(1), 133–152.

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.

Evans, T. (1968). A heuristic program to solve geometric analogy problems. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, MA: MIT Press.

Goel, A., & Bhatta, S. (2004). Use of design patterns in analogy-based design. *Advanced Engineering Informatics, 18*(2), 85–94.

Goel, A., Bhatta, S., & Stroulia, E. (1997). Kritik: An early case-based design system. In M. Maher & P. Pu (Eds.), *Issues and applications of case-based reasoning in design* (pp. 87–132). Mahwah, NJ: Erlbaum.

Goel, A., & Chandrasekaran, B. (1988). Integrating case-based and model-based reasoning for design problem solving. In *Proceedings of the AAAI-88 workshop on AI in design, Minneapolis, MN, August, 1988*.

Griffith, T., Nersessian, N., & Goel, A. (2000). Function-follows-form transformations in scientific problem solving. In *Proceedings of the*

twenty-second annual conference of the cognitive science society. Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's advanced progressive matrices. *Journal of Vision, 11*(10), 1–11, article no.10.

Hofstadter, D. (Ed.). (1995). *Fluid concepts and creative analogies*. New York: Basic Books.

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

Indurkhya, B. (1998). On creation of features and change of representation. *Cognitive Studies, 5*(2), 43–56.

Kirby, J. R., & Lawson, M. J. (1983). Effects of strategy training on progressive matrices performance. *Contemporary Educational Psychology, 8*, 127–140.

Kokinov, B. (1998). Analogy is like cognition: Dynamic, emergent, and context-sensitive. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences*. Sofia: NBU Press.

Kosslyn, S. M., Ball, T., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 47–60.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York, NY: Oxford University Press.

Kunda, M., & Goel, A. (2008). How Thinking in Pictures can explain many characteristic behaviors of autism. In *Proceedings of the 7th IEEE international conference on development and, learning* (pp. 304–309).

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. (2012). Reasoning on the Raven's advanced progressive matrices test with iconic visual representations. In *Proceedings of the 34th annual conference of the cognitive science society*.

Kunda, M., Soulières, I., Mottron, L., & Goel, A. K. (2011). Comparing patterns of errors on the Raven's progressive matrices test: Strategy differences among typically developing individuals, individuals with autism, and computational models. Presented at *the international meeting for autism research (IMFAR)*.

Leyton, M. (2001). *A generative theory of shape*. Berlin: Springer.

Lovett, A., Forbus, K., & Usher, J. (2010). A structure-mapping model of Raven's progressive matrices. In *Proceedings from the 32nd annual conference of the cognitive science society*.

Lynn, R., Allik, J., & Irwing, 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.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence–picture relationships. *Journal of Verbal Learning and Verbal Behavior, 17*, 493–507.

McGreggor, K., Kunda, M., & Goel, A. K. (2010). A fractal analogy approach to Raven's test of intelligence. In *Proceedings of the AAAI-2010 workshop on visual representations and reasoning, Atlanta, GA*.

Meo, M., Roberts, M. J., & Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's progressive matrices. *Intelligence, 35*, 359–368.

Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.

Neubauer, A. C., & Freudenthaler, H. H. (1994). Reaction times in a sentence–picture verification test and intelligence: Individual strategies and effects of extended practice. *Intelligence, 19*(2), 193–218.

Ojha, A., & Indurkhya, B. (2009). Perceptual vs. conceptual similarities and creation of new features in visual metaphor. In B. Kokinov, K. Holyoak, & D. Gentner (Eds.), *New frontiers in analogy research*. Sofia: New Bulgarian University Press.

Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory and Cognition, 6*(2), 91–97.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's progressive matrices test. *Cognitive Psychology, 33*(1), 43–63.

Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence, 30*, 41–70.

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science, 3*(1), 140–153.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence–picture verification. *Cognitive Psychology, 40*, 261–295.

Sahyoun, C., Soulières, I., Belliveau, J., Mottron, L., & Mody, M. (2009). Cognitive differences in pictorial reasoning between high-functioning autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders, 39*(7), 1014–1023.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*(1), 36–71.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*(2), 166–183.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. *Advances in the Psychology of Human Intelligence, 2*, 47–103.

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), 4082–4107.

Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition*. London: Macmillan.

Stafford, B. (2001). *Visual analogy: Consciousness as the art of connecting*. Cambridge, MA: MIT Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

Tversky, B. (1975). Pictorial encoding of sentences in sentence–picture comparison. *Quarterly Journal of Experimental Psychology, 27*, 405–410.

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.

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence, 34*, 261–272.

Villardita, C. (1985). Raven's colored progressive matrices and intellectual impairment in patients with focal brain damage. *Cortex, 21*(4), 627–634.

Yaner, P., & Goel, A. (2006). Visual analogy: Viewing retrieval and mapping as constraint satisfaction. *Journal of Applied Intelligence, 25*(1), 91–105.

Zitová, B., & Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing, 21*, 977–1000.