Addressing the Raven's Progressive Matrices Test of "General" Intelligence

Maithilee Kunda, Keith McGreggor, & Ashok Goel

Design & Intelligence Laboratory, School of Interactive Computing
Georgia Institute of Technology, Atlanta, Georgia 30332, USA
mkunda@gatech.edu, keith.mcgreggor@gatech.edu, goel@cc.gatech.edu

Abstract

The Raven's Progressive Matrices (RPM) test is a commonly used test of general human intelligence. The RPM is somewhat unique as a general intelligence test in that it focuses on visual problem solving, and in particular, on visual similarity and analogy. We are developing a small set of methods for problem solving in the RPM which use propositional, imagistic, and multimodal representations, respectively, to investigate how different representations can contribute to visual problem solving and how the effects of their use might emerge in behavior.

Introduction

The Raven's Progressive Matrices (RPM) test[1] is a standardized intelligence test that consists of visually presented, geometric-analogy-like problems in which a matrix of geometric figures is presented with one entry missing, and the correct missing entry must be selected from a set of answer choices. Figure 1 shows an example of a problem that is similar to one of the problems in the Standard Progressive Matrices (SPM).

Although the test is supposed to measure only eductive ability, or the ability to extract and understand information from a complex situation (Raven, Raven, & Court 1998), the RPM's high level of correlation with other multi-domain intelligence tests have given it a position of centrality in the space of psychometric measures (Snow, Kyllonen, & Marshalek 1984), and it is therefore often used as a test of general intelligence. Using the RPM as a measure of general intelligence, though it consists only of problems in a single, nonverbal format, stands in contrast to using broader tests like the Wechsler scales, which are comprised of subtests across several different verbal and nonverbal domains.

Despite its widespread use, neither the computational nor the cognitive characteristics of the process of solving the RPM are well understood. Hunt (1974) gives a theoretical account of the information processing demands of certain problems from the Advanced Progressive Matrices (APM), in which he proposes two qualitatively different solution algorithms—"Gestalt," which uses visual representations and perceptually based operations, and "Analytic," which uses feature-based representations and logical operations—that could yield identical results on at least portions of the test.

Our work expands on Hunt's idea by asking whether qualitatively different systems of representation can lead to identical performance on the RPM. Our question is theoretically interesting for the study of cognition and AI because the fact that the RPM correlates so well with broader tests of intelligence suggests that the specific information processing capacities tapped by the RPM may

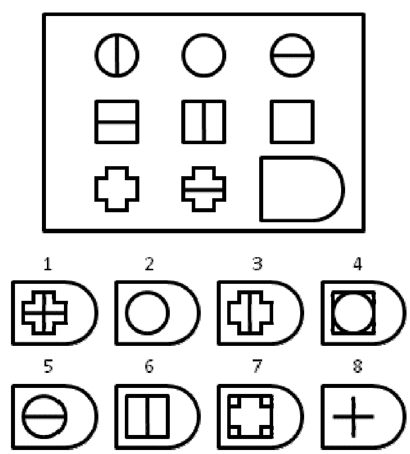

Figure 1. Example problem similar to one in the Standard Progressive Matrices (SPM) test.

[1] A note on the Raven's family of tests: We use RPM to denote the Raven's Progressive Matrices paradigm without referring to any particular version of the test. Specific versions that we discuss include the original Standard Progressive Matrices (SPM) and the Advanced Progressive Matrices (APM), which was developed as a more difficult test to reduce the ceiling effects sometimes found with the SPM (Raven, Raven, & Court 1998). Colored Progressive Matrices (CPM) is a simpler version of the test often used with children or other less mentally able individuals; we do not address the CPM in this work.

be central to domain-general processes of reasoning used across a variety of cognitive tasks. If different schemes of representation can provide equivalent performance, then this raises the issues of 1) whether these broader reasoning processes can (or must) also be instantiated with different types of representations, 2) to what extent a single individual may (or must) draw on these different representations for reasoning tasks, and 3) to what degree individual differences may (or must) result from variations in the underlying representations being used.

Our central question is also of immense practical import because the RPM family of tests is used extensively in clinical, educational, vocational, and scientific settings as an accurate assessment of intelligence. Therefore, interpretations of RPM scores should be made only with a thorough understanding of what cognitive implications these scores do and do not provide.

Evidence for Different RPM Strategies

While within- and between-individual differences on the RPM are generally treated as being those of degree and not of type, there is evidence that the use of qualitatively different solution strategies may be responsible for some of these differences. Within-individual differences have been linked to the presence of different problem types on the RPM, i.e. factor analyses and other behavioral markers have suggested that different problems may admit distinct solution strategies (DeShon, Chan, & Weissbein 1995; Dillon, Pohlmann, & Lohman 1981).

Recent studies have highlighted the possibility of there being qualitative differences in RPM solution strategies between individuals with autism and "neurotypical" individuals. Autism is a developmental disorder marked by significant language delay and atypical behaviors in social interaction, communication, and stereotyped or repetitive patterns of behavior and interest (*DSM-IV-TR* 2000). While individuals with autism show impaired performance on many cognitive tasks, many studies report intact or superior performance on certain visuospatial tasks (e.g. Joliffe & Baron-Cohen 1997).

Studies of intelligence tests have shown that, unlike neurotypical individuals who show correlated scores, individuals with autism often demonstrate RPM scores that are much higher than their Wechsler scores (Mottron 2004; Dawson et al. 2007). Individuals with Asperger's syndrome show a similar pattern (Hayashi et al. 2008).

One possible explanation for these results is that individuals with autism might be predisposed towards reasoning visually (Kunda & Goel 2008a, 2008b). In this case, they might find the RPM amenable to a visual reasoning solution but the verbal Wechsler subtests very difficult. Recent neuroimaging evidence is consistent with this possibility (Soulières et al. 2009).

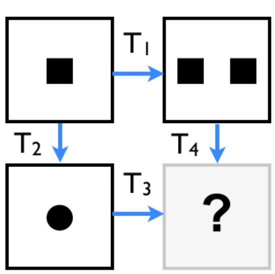

Figure 2. Illustration of four image transformations implicit in a two-by-two RPM problem matrix.

Building Computational Accounts of the RPM

Existing computational accounts of problem solving on the RPM, with the exception of Hunt's early work in 1974, use propositional representations and assume that individual differences result from quantitative differences in the problem solving architecture, e.g. number of goals that can be managed, types of shapes that can be discriminated, etc. (e.g. Carpenter, Just, & Shell 1990; Lovett, Forbus, & Usher 2007). In contrast, we are developing three methods for solving RPM problems, each of which uses a fundamentally different representational substrate.

The first is a purely imagistic method that relies on fractal encodings of an RPM matrix. The fractal representations use only grayscale pixel values from an image of an RPM problem and are mathematical abstractions quite rigorously grounded in the theory of fractal image compression (Barnsley & Hurd 1992).

The second method uses a multimodal representation in which elements from an RPM problem are represented using a visuospatial-symbolic scheme, in which the elements can be manipulated symbolically but only with operations based in visual perception, e.g. translation, rotation, scaling, etc. This representation follows the work of Kosslyn, Thompson, and Ganis (2006) on depictive mental representations and is similar to representations used in our earlier work on visual analogy (Davies & Goel 2001; Davies, Goel, & Yaner 2008).

The third method uses propositional representations similar to those in Carpenter, Just, and Shell (1990). We view this chiefly as a control representation with which to compare our first and second methods. We now present our first and second methods in greater detail.

Fractal Representations and the RPM

An RPM problem can be viewed as a sequence of images (ordered in rows and columns), where some unknown transformation T can be said to transform one image into a corresponding adjacent image. In a typical two-by-two RPM problem, there are four such transformations, as shown in Figure 2. (RPM problems can also have three-by-three matrices, which we do not address in this paper.)

RPM problems are formulated to suggest that these transformations are pairwise analogous (i.e. the two row transformations are analogous to one another). We seek to solve a problem by determining which of the candidate solutions yields the most analogous transformations. To this end, we explored the fractal encoding of one image in terms of another as a representational basis for calculating and discovering the underlying analogies.

The mathematical derivation of fractal encoding expressly depends upon the notion of real world images, i.e. images that are two dimensional and continuous. A key observation is that all naturally occurring images we perceive appear to have similar, repeating patterns. Another observation is that no matter how closely you examine the real world, you find instances of similar structures and repeating patterns.

This suggests that it is possible to describe the real world in terms other than those of shapes or traditional graphical elements—in particular, terms which capture the similarity and repetition alone.

The theorem at the heart of the fractal encoding algorithm can be stated concisely:

> *For any particular real world image D, there exists a finite set of affine transformations T which, if applied repeatedly and indefinitely to any other real world image S, will result in the convergence of S into D.*

The Fractal Encoding Algorithm

Given an image D, the fractal encoding algorithm seeks to discover the set of transformations T. The algorithm is considered "fractal" for two reasons: first, the affine transformations are generally contractive, which leads to convergence, and second, the convergence of S into D can be shown to be the mathematical equivalent of considering D to be a strange attractor.

Here are the general steps of the algorithm for encoding an image D in terms of another image S:

1) Decompose D into a set of N smaller images $\{d_1, d_2, d_3, ..., d_n\}$. These individual images are sets of points.
2) For each image d_i:
    a) Examine the entire source image S for an equivalent image s_i such that an affine transformation of s_i will result in d_i. This affine transformation will be a 3x3 matrix, as the points within s_i and d_i under consideration can be represented as the 3-D vector $<x, y, c>$ where c is the (grayscale) color of the 2-D point $<x,y>$.
    b) Let T_i represent the compact representation of the discovered affine transformation.
3) The set $T = \{T_1, T_2, T_3, ..., T_n\}$ is the fractal encoding of the image D.

The fractal transforms we construct are sets of specific affine transformations which describe the alteration and colorization of fragments of the source image that will collage into the destination image. These fragments are generally blocks of pixel data inferior in size to the whole image. The fractal encoding indicates compactly 1) from where the source material originates and 2) how to manipulate it geometrically (copy, rotation, or flip) and photometrically (altering the block's luminosity).

While it is tempting to treat contiguous subsets of these transformations as features, note that their derivation does not follow strictly Cartesian notions (e.g. adjacent material in the destination might arise from strongly non-adjacent source material). With this in mind, we consider (in our present implementations) each of these block-level transformations to be independent of one another, and we only construct candidate fractal features for matching from single block-level transformations. Each such transform yields a very small finite set of fractal features.

We generate fractal solutions to RPM problems by examining all possible pairwise transforms and calculating a measure of similarity for each pair. This metric reflects similarity as a comparison of the number of fractal features shared between candidate pairs taken in contrast to the joint number of fractal features found in each pair member (Tversky 1977). The solution is chosen as the answer that results in the highest measured similarity for both row and column pairs in the RPM problem.

Affine Symbolic Reasoning and the RPM

As described at the start of the fractal section above, pairs of images in a two-by-two RPM problem can be viewed as having some unknown transformations T that apply either across a row or down a column. Following the view of mental imagery as supporting operations on depictive representations (Kosslyn, Thompson, & Ganis 2006), we designed our second method around affine transformations (such as rotation and reflection) and other depictive operations such as image addition and subtraction (and so we call this method "affine-extended").

In particular, this method seeks to explicitly characterize the transformation across any given row or column of an RPM problem matrix as being one of (or a composition of) these types of transformations. Unlike the fractal method, the affine method then applies this transformation to the row or column with the empty entry to generate a predicted value for the missing image.

Only after a prediction has been made does the algorithm examine the set of answer choices. It chooses an answer based on some measure of visual similarity, for instance choosing the answer that minimizes the sum-squared-difference between the intensity (grayscale) values of the predicted image and the answer image.

The Affine-Extended Algorithm

The affine-extended algorithm seeks to discover a single transformation T that holds across any of the complete rows or columns of the RPM problem matrix. In particular, the algorithm has access to a memory store containing a set of basic transformations along with a finite subset of possible compositions of such transformations, from which it can draw specific candidate transformations to test on the matrix. The algorithm proceeds as follows, in the case of a two-by-two RPM matrix problem:

1) For each transformation T_i in memory:
   a) For the top row, check to see if T_i holds.
      i) If so, apply T_i to the bottom row to generate a guess for the missing image. Go to step 2.
      ii) If not, continue to step 1b.
   b) For the left column, check to see if T_i holds.
      i) If so, apply T_i to the right column to generate a guess for the missing image. Go to step 2.
      ii) If not, repeat step 1 with the next T_i.
      iii) If there are no more transformations in memory, then the algorithm halts.
2) For each answer choice A_i:
   a) Compute the sum-squared-difference (SSD) between the predicted image and A_i.
3) Choose the answer choice A_i that minimizes the SSD value.

Unlike the fractal method, the affine-extended method will not always produce a solution. The memory set is restricted to a finite set of possible relations in order to prevent the algorithm from stalling indefinitely on a single problem. Despite this limitation, we hypothesize that the affine-extended method will in fact be able to solve a large fraction of the SPM problems.

Ambiguity and the Affine-Extended Method

Notice, in the algorithm above, that there are two biases in searching for the correct transformation. First, there is a bias in examining rows first and then columns. Second, there is a bias in the order in which transformations are retrieved from memory.

These biases would not matter if there were exactly one transformation from memory that held for a given matrix problem. However, the presence of ambiguity in some problems means that, if two or more transformations hold, or if a transformation holds for both rows and columns, the algorithm will use the first one that crosses its path, which may or may not be the correct one.

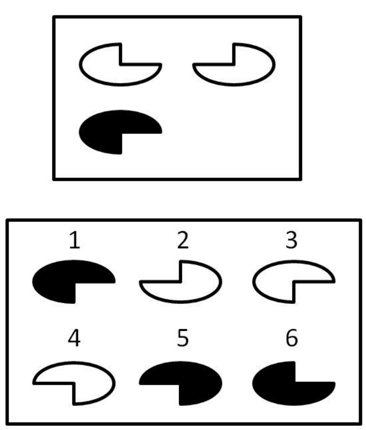

Figure 3. Example problem similar to one in the Standard Progressive Matrices (SPM), illustrating ambiguity in possible solutions when using the affine-extended method.

Figure 3 gives an example of a problem with this kind of ambiguity; this problem is analogous to one in the actual SPM test. One transformation that could hold in this problem would be, across the top row, reflection about a vertical axis. Using this transformation on the bottom row, the predicted answer would be identical to answer #5.

However, one could also have a transformation across the top row that consists of a rotation 90° to the left and then a scaling along the vertical axis (i.e. "squishing" the figure down vertically). Applying this transformation to the bottom row would predict an image identical to answer #6. So which is correct?

There are several ways, which we are currently investigating, in which this conflict could be resolved. First, the algorithm could just stick with its first prediction, and sometimes its biases would result in the correct answer being chosen, and sometimes not. Another option is to explicitly assign biases to the transformations in memory, where, for example, a transformation comprised of a single operation (e.g. reflection) would be examined before transformations comprised of multiple operations (e.g. rotation plus scaling). This approach puts additional requirements on the memory store and assumes that such rankings of transformations are somehow justified.

A third approach is to add an additional step to the algorithm that examines the final matrix (with the predicted answer choice in place) for qualities of symmetry or other global measures. If symmetry were the measure, then the algorithm would choose answer #5 over answer #6 in the example problem shown in Figure 3.

This approach raises questions as to what precisely the RPM tests are measuring, if such problem ambiguities are resolved by a fixed, global scale of optimal answers. It could be that aspects of symmetry play a key role, in which case it would be interesting to compare the demands of computing symmetry on the RPM to the requirements for recognizing symmetry in other cognitive tasks.

Contrasts with Propositional Representations

Carpenter, Just, and Shell (1990) describe a computational model that simulates solving RPM problems using propositional representations. Their model is based on the traditional production system architecture, with a long-term memory containing a set of productions and a working memory containing the current state of problem solving (e.g. current goals). Productions are based on the relations among the entities in a RPM problem, for example, the location of the dark component in a row, which might be the top half in the top row of a problem, bottom-half in the bottom row, and so on.

Their computer model is able to emulate human problem solving on most RPM problems (from the Advanced Progressive Matrices test) up to the performance levels of the highest scoring subjects in their study. However, as described in their paper, the model "operates on a hand-coded, symbolic description of each matrix entry." Thus, visual information is encoded implicitly in their coded matrix entries and also in the set of production rules but is nowhere explicitly accessed or used by the model.

Preliminary Results

In this section, we will describe solutions based upon each of our two novel computational methods.

Example Using the Fractal Method

As an example, we shall use the "arrow" problem shown in Figure 4. Also, while the complete fractal method examines each of the problem's analogies, we shall restrict this detailed discussion to just one of these transformations (T_1, as labeled in Figure 2).

The initial transformation T_1 is the fractal encoding of the transformation from the upper left arrow into the upper right arrow. When encoded using a block size of 32 x 32 pixels, this encoding generates 39 distinct fractal features.

Each candidate answer is encoded likewise, from the upper left arrow into the candidate, each resulting in between 27 and 45 distinct features. On the determination of the set of fractal features for each candidate, a measure of similarity S between the candidate transform C and the target transform T_1 is calculated using the following formula, also known as the ratio model (Tversky 1977):

$$S(T,C) = f(T \cap C) \,/\, [\, f(T \cap C) + \alpha f(T - C) + \beta f(C - T) \,]$$

where $f(A)$ is the number of features in the set A. For our initial work, we have chosen values of $\alpha = \beta = 1.0$, which, according to Tversky, results in this simplification:

$$S(T,C) = f(T \cap C) \,/\, f(T \cup C)$$

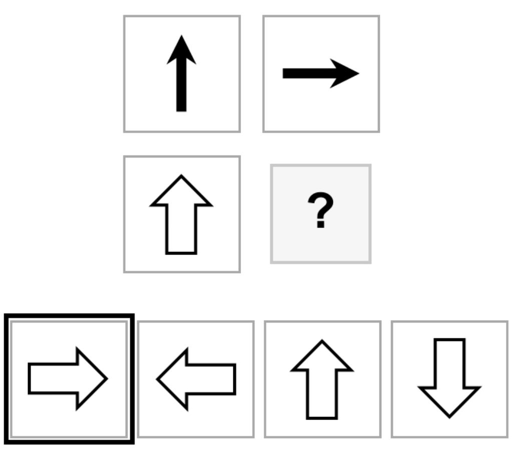

Figure 4. Example problem for illustrating the execution of the fractal and affine-extended methods.

The answer with the highest calculated similarity is deemed correct. For the arrow problem, using a 32 x 32 block size, the similarity measures for each answer are:

$$S(T,C_1) = 21 \,/\, (21+18+24) \cong 0.333333$$
$$S(T,C_2) = 15 \,/\, (15+24+30) \cong 0.217391$$
$$S(T,C_3) = 16 \,/\, (16+23+11) \cong 0.32$$
$$S(T,C_4) = 14 \,/\, (14+25+29) \cong 0.205882$$

Therefore, the fractal method chooses as its answer #1.

Example Using the Affine-Extended Method

For this example, we used a limited version of the affine-extended method that holds in memory the following image transformations, in this order:

1) identity
2) horizontal reflection
3) vertical reflection
4) rotate 90°
5) rotate 180°
6) rotate 270°

These transformations are successively tested against the images in the top row and then left column to see if they hold. To check each transformation, the first entry in the row or column is transformed accordingly and compared to the second entry using the sum-squared-difference (SSD) of the pixel intensity values. A threshold value determines whether the transformation holds, and the algorithm stops testing transformations once a valid one has been found.

For the example shown in Figure 4, the SSD values for each of the first three transformations were around 150 million, which fell well above the threshold, and the SSD for the fourth transformation was exactly zero (due to the noiselessness of the input images). So, the fourth transformation was chosen as the correct one.

Once a valid transformation is found, the first entry in the remaining row or column is transformed according to this rule and compared to the answer choices, again using the sum-squared difference of the pixel values. The answer with minimum difference is chosen as correct.

In this example problem, the SSD values calculated for each answer choice were as follows:

1) 0.0
2) 3.19×10^8
3) 3.57×10^8
4) 3.54×10^8

Thus, the affine-extended method also chooses as its answer #1.

Discussion

Presently, we are engaged in defining, implementing, and testing our three systems of representation on subsets of the RPM. We expect to answer these questions concerning accuracy, efficiency, and behavioral effects:

- Can these different methods produce the same (correct) behavior on RPM problems, and if so, on which of the problems and why? Are there intrinsic aspects of the various problems that make them amenable to solution using particular methods?
- Do these different methods confer processing speed advantages for certain RPM problems?
- What behavioral markers can we determine to distinguish among these three methods, and given these markers, how might we test for their presence in human subjects?

With the answers to these questions in hand, we propose to have established the first firm cognitive and computational account of the Raven's Progressive Matrices test using multiple-representation methods.

Acknowledgments

This research has been partially supported by an NSF grant (IIS Award #0534266) entitled "Multimodal Case-Based Reasoning in Modeling and Design;" by the Office of Naval Research through an NDSEG fellowship; and by the NSF GRFP fellowship program. We also thank Gregory Abowd for his encouragement of this work.

References

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders, Revised 4th Ed.* Washington, DC: American Psychiatric Association.

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

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.

Davies, J., & Goel, A.K. 2001. Visual analogy in problem solving. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, 377-382. Seattle, WA: Morgan Kaufmann.

Davies, J., Goel, A.K, & Yaner, P.W. 2008. Proteus: A theory of visual analogies in problem solving. *Knowledge-Based Systems* 21 (7): 636-654.

Dawson, M., Soulières, I., Gernsbacher, M. A., & Mottron, L. 2007. The level and nature of autistic intelligence. *Psychological Science* 18 (8): 657-662.

DeShon, R. P., Chan, D., & Weissbein, D. A. 1995. Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence* 21 (2): 135-155.

Dillon, R. F., Pohlmann, J. T., & Lohman, D. F. 1981. A factor analysis of Raven's Advanced Progressive Matrices freed of difficulty factors. *Educational and Psychological Measurement* 41: 1295–1302.

Hayashi, M., Kato, M., Igarashi, K., & Kashima, H. 2008. Superior fluid intelligence in children with Asperger's disorder. *Brain and Cognition* 66 (3): 306-310.

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

Jolliffe, T., & Baron-Cohen, S. 1997. Are people with autism and Asperger syndrome faster than normal on the embedded figures test? *Journal of Child Psychology and Psychiatry and Allied Disciplines* 38 (5): 527-534.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. 2006. *The Case for Mental Imagery*. New York, NY: Oxford University Press.

Kunda, M., & Goel, A.K. 2008a. How Thinking in Pictures can explain many characteristic behaviors of autism. In *Proceedings of the 7th IEEE International Conference on Development and Learning*, 304-309. Monterrey, CA.

Kunda, M., & Goel, A.K. 2008b. What can pictorial representations reveal about the cognitive characteristics of autism? In Stapleton, G., Howse, J., and Lee, J. eds. *Proceedings of the 5th International Conference on the Theory and Application of Diagrams, LNAI* 5223: 103-117.

Lovett, A., Forbus, K., & Usher, J. 2007. Analogy with qualitative spatial representations can simulate solving Raven's Progressive Matrices. In *Proceedings from the 29th Annual Conference of the Cognitive Science Society*, 449-454. Nashville, TN: Cognitive Science Society, Inc.

Mottron, L. 2004. Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders* 34: 19-27.

Raven, J., Raven, J. C., & Court, J. H. 1998. *Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 1: General Overview*. San Antonio, TX: Harcourt Assessment.

Soulières, I., Dawson, M., Samson, F., Barbeau, E. B., Sahyoun, C., Strangman, G. E., et al. 2009. Enhanced visual processing contributes to matrix reasoning in autism. *Human Brain Mapping*. Forthcoming.

Snow, R., Kyllonen, P., & Marshalek, B. 1984. The topography of ability and learning correlations. In Sternberg, R. ed. *Advances in the Psychology of Human Intelligence* 2: 47-103. Hillsdale, NJ: Erlbaum.

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