Gentner, D., & Sagi, E. (2006). Does "different" imply a difference? A comparison of two tasks. In R. Sun & N. Miyake (Eds.). Proceedings of the Twenty-eighth Annual Meeting of the Cognitive Science Society, 261-266.

Does "Different" Imply a Difference? A Comparison of Two Tasks

Dedre Gentner (gentner@northwestern.edu)
Eyal Sagi (ermon@northwestern.edu)
Department of Psychology, Northwestern University
2029 Sheridan Road, Evanston, IL 60208 USA

Abstract

One of the most interesting predictions of structure-mapping theory (Gentner, 1983) is that differences are more easily identified when the comparison involves stimuli that are easily aligned. Evidence for this claim comes from studies in which participants state differences between stimuli pairs (e.g. Gentner & Markman, 1994). These results are at odds with results from tasks in which participants are asked to determine whether pairs of images differ or not. In such tasks, it is often found that participants are faster to make such a determination when the images differ than when they are similar (Luce, 1986). However, comparing these results is difficult because the two lines of research employ different experimental designs and methodologies. This paper describes two experiments that contrast the two results within the same framework in an attempt to examine more closely the differences between the tasks.

Keywords: Comparison, Perception, Similarity

Introduction

According to structure-mapping theory (Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983, 2003; Gentner & Markman, 1993, 1994; Markman & Gentner, 1993) the process of comparison involves the *alignment* of the two representations compared. The alignment of two representations goes beyond the identification of shared features; it also requires finding correspondences between the *relations* that connect the features.

Because structure-mapping postulates that similarity involves relational as well as featural correspondences, it can predict the discovery of alignable differences: features that differ between the stimuli but occupy a similar position in the structure. For example, a structural alignment between a bird and a person might involve a correspondence between the bird's wing and the person's hand, despite the closer local resemblance between the bird's foot and the person's hand. Once an alignment has been established, differences between the objects being compared are easily spotted. According to structure-mapping, commonalities and differences that are connected to the common structure are more salient than those that are not. This means that alignable differences are more salient than non-alignable differences. This makes intuitive sense, for it leads to a focus on differences that are relevant to the common causal or perceptual structure that is the focus of the comparison. But it leads to the rather paradoxical prediction that in general, there will be more salient differences for high-similar than for low-similar pairs (because in general, high-similar pairs have larger common systems and thus more slots for alignable differences). These claims concerning the relation between differences and commonalities distinguish structure-mapping from other theories of similarity such as feature-set intersection models and mental distance models, as amplified below.

The prediction that differences are easier to detect for high-similar pairs than for low-similar pairs was borne out in a speeded difference task. Gentner and Markman (1994) gave participants the task of finding one difference between as many word pairs as possible in a rather brief time period. Participants identified differences for many more high-similar pairs than low-similar pairs. Gentner and Gunn (2001) asked people to compare word pairs and write a commonality, and then gave them a speeded-difference task. Participants generated more differences (mostly alignable differences) for the previously compared pairs than for new pairs, showing the close connection between alignment and difference-noticing. More relevant here, participants also generated more differences for high-similar pairs than for low-similar pairs.

In a test of this framework for perceptual comparison, Markman and Gentner (1996) gave participants image pairs and asked them to list either differences or commonalities. Again, participants listed more differences for highly similar images than for less similar ones. These findings are consistent with the structure-mapping claim that participants will find it easier to note differences between concepts and images that are fairly similar (and consequently more alignable) than between concepts and images that are substantially different (and therefore difficult to align) (e.g., Kurtz, Miao, & Gentner, 2001).

Same-different judgments

On the face of it, the above findings seem at odds with a venerable body of research on tasks involving same-different judgments. A well-established result is that the more similar two images are, the more difficult it is to identify that they are different (e.g. Farell, 1985; Goldstone & Medin, 1994; Luce, 1986; Posner & Mitchell, 1967; Tversky, 1969). That is, the more similar two things are, the longer people require to say "different" and the more likely they are to erroneously identify the pair as "same." This result runs in the opposite direction from the previously described findings in which participants found it easier to identify differences in similar images than in dissimilar ones.

A possible resolution would be to assume that alignment is crucial in the identification of a difference between two images, but plays no role (or only a minor role) in the decision as to whether the two images are different. This could come about because the two tasks call on two distinct similarity processes. Alternatively, it could be that the two tasks tap into different stages of the same process. This second possibility fits with a proposal by Markman and Gentner (2005): that the local-to-global alignment process postulated in structure-mapping can yield two kinds of output: one, based on a full structural alignment, and the other based on a readout from the initial parallel matching step. The full process is relatively slow, and provides a specific alignment: a common structure, and typically some alignable differences and candidate inferences. In contrast, the fast early readout gives rise to estimates of overall similarity without taking structure into account.

Our goal in this research is to compare these two distinct tasks—a same-different judgment and a difference-identification task. In order to make such a comparison possible, the two tasks had to be made as similar as possible. Same-difference tasks typically use reaction times and error rates as dependant measures. Therefore, we adapted these measures for an identification-of-differences task. As discussed below, Experiment 1 employs reaction time as the primary measure for both tasks.

In Experiment 1 we designed a highly controlled set of stimuli wherein similar pairs differ on a single, salient, feature while dissimilar pairs differ on a multitude of salient features (see Figure 1)[1]. According to many models of comparison, this should make the task of identifying a difference more difficult. For instance, mental distance models (e.g., Nosofsky, 1986; Shepard, 1974; Shoben, 1983) model similarity in terms of the distance between points within a multi-dimensional mental space. Their relative position within this space can then be used to determine what relevant differences (and similarities) exist between the objects, as well as measure how different the two objects are from one another. Generally speaking, the farther apart two points are within the space, the easier it should be both to detect *that* they are different and to find specific differences (difference in dimensional values) between them. For example, the top two images in Figure 1 are very similar and therefore the mental distance between them should be quite small. The same is true for the bottom two images. In contrast, the distance between the two bottom images and the ones on top would be much greater because of the many differences that exist between them.

In feature-intersection models (e.g. Tversky, 1977), objects are represented by sets of independent features. Comparison is based on the examination of those sets, extracting shared and distinct features. Similarity is increased by shared features and decreased by distinctive features. The reverse applies when computing a difference judgment; in addition, people are assumed to weight distinctive features more heavily for difference judgments. The greater the number of distinctive features, the easier it should be both to detect *that* two objects are different and to find distinctive features. For instance, the top-left image in Figure 1 varies from the top-right one on a single feature – the color of the innermost circle. However, when comparing the top-left image with the bottom-left one, almost any feature becomes a distinctive feature. According to feature-interaction models, it should therefore be much easier to tell the top-left image from the bottom-left one than from the top-right one. The sheer number of distinctive features should also make it easier to identify a difference between the two leftmost images than between the two topmost images.

Both mental distance models and feature models would therefore predict a positive relation between the two tasks: the fewer the differences that exist between two objects, the harder it should be *both* to detect that they are different and to identify a specific difference between them.

Structure-mapping theory makes a different prediction. Because the similar images (e.g., shields A and B in Figure 1) share many features as well as a common organizing structure (and the shared features are structurally consistent with the shared structure), they should be highly alignable. In contrast, images A and C should be difficult to align, because of their low degree of structural overlap. Structure-mapping theory would therefore predict that participants will find it easier to identify a specific difference between A and B than between A and C.

Experiment 1

Following structure-mapping theory, we predicted that responses in the difference-identification task should take longer than those in the same-different task. Further, based on Markman and Gentner's two-phase conjecture, (a) participants in the difference-identification task should be faster to respond to a similar pair than to a dissimilar one; whereas (b) participants in the same-different task should be faster to make a "different" judgment for a dissimilar pair than for a similar one.

Method

Participants The participants were undergraduate students at Northwestern University who participated for class credit; 24 in the same-different condition and 20 in the difference-identification condition.

Materials The materials for this experiment were 60 images designed in the likeness of Heraldic shields. Forty of the images were pairs of highly similar and alignable images. The difference between the two images in such a pair was in a single design element (e.g. the crest, a background pattern, etc.). These 20 pairs were then combined into groups of 2 pairs, such that the images of one pair would be highly dissimilar to the images of the other pair (see Figure 1). Each such group was then arranged in two possible

[1] A word on terminology is in order. We follow the conventions of field in contrasting high-similar pairs with low-similar pairs. Analogy theory makes a finer set of distinctions, as discussed below, but to link with the field, we use the traditional dichotomy for these studies.

arrangements of image pairs – one arrangement consisting of two pairs that were each similar (high-sim pairs), and the other consisting of two pairs that were dissimilar (low-sim pairs). The remaining 20 images were used to create 20 pairs of identical images ('same' pairs).

Each participant saw 10 high-sim pairs and 10 low-sim pairs. Additionally, participants in the same-different condition saw 20 'same' pairs, so that only half of the images viewed by these participants were different.

Finally, 10 pairs consisting of arrangements of geometrical forms were used for training. Of these 10 pairs, 5 were identical pairs and 5 were non-identical pairs. The 5 identical pairs were only presented to participants in the same-different condition.

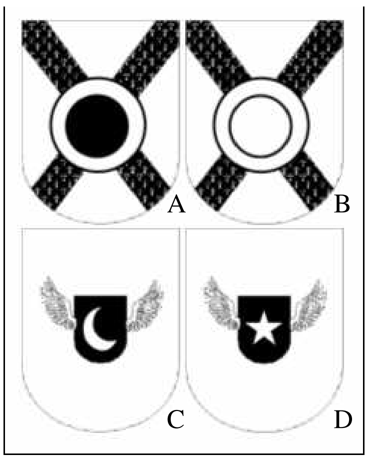

Figure 1: Sample stimuli from experiment 1. Images in the same row represent high-sim pairs; images in the same column represent low-sim pairs

Procedure Participants were seated in front of a computer and presented with the instructions on the computer screen. After reading the instructions participants completed the training phase. Following the training phase, participants were presented with the experimental image pairs. These images were presented in two blocks of equal length. The presentation of each pair was preceded by a half-second fixation period during which a crosshair appeared at the center of the screen.

In the same-different condition, participants were instructed to indicate whether each pair consisted of identical images or non-identical images. Participants indicated their decision by pressing the left- or right- control key on the computer keyboard. Left-right position was reversed for half the participants. The time between the onset of presentation of an image pair and a participants' response was recorded.

In the different-identification condition, participants were instructed to press the space key after identifying a difference between each presented image pair. Following each such presentation, participants then typed a specific difference between the images. As in the same-different condition, the time between the onset of presentation of an image pair and the pressing of the space key was recorded.

Results and discussion

The mean results are shown in Figure 2. As predicted, participants were faster to make a 'different' judgment for low-sim pairs (*M* = 0.95sec) than for high-sim pairs (*M* = 1.36sec), but were slower to identify a difference between low-sim pairs (*M* = 9.26sec) than between high-sim pairs (*M* = 6.73sec). The average response time for each of the two types of experimental image pairs (high-sim and low-sim) was calculated for each participant and the results were analyzed using a repeated-measures ANOVA of Task x Similarity. This analysis revealed a highly reliable interaction between the between-s variable of task (same-different judgment vs. difference-identification) and the within-s variable of similarity (high-sim vs. low-sim). This interaction is shown in Figure 2, $F(1,42) = 17.00$, $MS_e = 47.41$, $p < .001$.

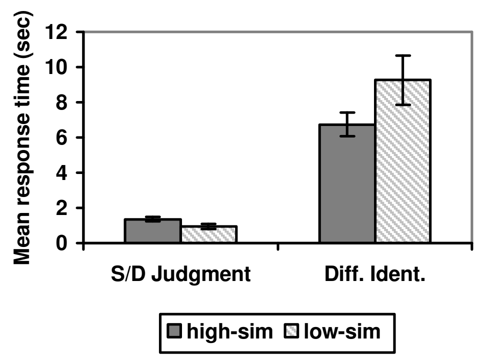

Figure 2: Results from experiment 1 (error bars represent the standard error of the mean)

Additionally, there were also reliable main effects for both similarity ($F[1,42] = 8.7$, $MS_e = 24.26$, $p < .01$) and task ($F[1,42] = 84.77$, $MS_e = 1020.52$, $p < .001$). Importantly, planned contrasts for both tasks revealed that the observed performance differences in response time across similarity levels were statistically reliable in both cases (same-different judgment – $F[1,23] = 133.44$, $MS_e = 2.11$, $p < .001$; difference-identification – $F[1,19] = 10.4$, $MS_e = 63.93$, $p < .01$).

An item ANOVA revealed similar results – a reliable interaction ($F[1,76] = 6.62$, $MS_e = 44.58$, $p < .05$) and a reliable main effect of task ($F[1,76] = 204.24$, $MS_e = 1375.18$, $p < .001$). However, the item analysis revealed only a marginally significant main effect of image-pair type ($F[1,76] = 3.29$, $MS_e = 22.16$, $p < .1$).

The results of Experiment 1 bear out the experimental hypothesis—that two different processes or two stages of the same process are involved in the two tasks. Participants were faster to distinguish two images when they were dissimilar, but slower to identify a specific difference between them. The vastly different mean response times are also consistent with this proposal: same-different judgments required only about 1.16 seconds, while identifying a difference required 8 seconds. Of course, the difference-identification times may have included some internal verbalization. Still, the difference is suggestive of different processes or stages.

Experiment 2

Experiment 1 demonstrated that the differences observed in the literature are congruent with the hypothesis Participants were faster to identify specific differences between similar (and alignable) images. In contrast, participants that were asked to judge whether two images were different or not were faster for dissimilar (difficult to align) images.

However, before embracing this possibility we need to ask whether it would hold for more naturalistic materials. The stimuli used in Experiment 1 were artificial and bore limited resemblance to images and stimuli participants are likely to encounter in the real world. Furthermore, because of the limited number of features and differences between the images, it is possible that participants elected to represent and compare the images in a strategic manner that was different than what they would normally use.

Experiment 2 aimed to replicate the results of Experiment 1 using stimuli that are less artificial and more akin to real-world images. We used sketches of plants taken from the Dover series (Harter, 1998). These stimuli differ in several ways from the artificially generated heraldry-like images in Experiment 1. First, as just noted, they are considerably more complex and variable. Second, we expect that in encoding the plant images in Experiment 2, participants will bring to bear more real-world knowledge. For instance, the identification of an image as a flower contributes to the identification of parts of it as petals, whereas the same parts might be identified as leaves in an image of a bush.

Additionally, as discussed above, the response patterns in Experiment 1 showed that participants took much more time comparing images in the difference-identification task than in the same-different task. Although this may point to a deep difference between the processes, we must also consider the possibility that the difference stems from task demands instead of a difference in processes. Perhaps participants in the same-different condition were satisfied with making intuitive snap judgments, whereas participants in the difference-identification condition wanted to make sure of the difference they identified. In order to minimize such effects, in Experiment 2 we presented the images for a short period of time (1500ms). However, in order to maintain the flow of the experiment, this form of presentation requires a slight change in the response time measure for the difference-identification task – instead of pressing 'space' after identifying the difference, participants are simply asked to type it in, and the measure used is the time they take before they start typing.

Because of the large effect size observed for the same-different condition in Experiment 1, we decided that fewer participants were needed in that condition in Experiment 2.

Method

Participants The participants were undergraduate students at Northwestern University who participated for class credit; 11 in the same-different condition and 40 in the difference-identification condition.

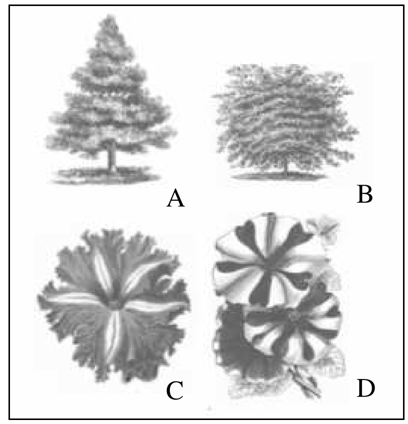

Figure 3: Sample stimuli from experiment 2. Images in the same row represent high-sim pairs; images in the same column represent low-sim pairs

Materials The stimuli for this experiment were 60 detailed drawing of plants. The arrangement of pairs was as for Experiment 1. Forty drawings were used for the experimental stimuli. As shown in Figure 3, each drawing belonged to both a similar (high-sim) pair, and a dissimilar (low-sim) pair. Participants saw each drawing only once (in either a high-sim pair or a low-sim pair). Twenty additional drawings were used to create 20 'same' pairs. Each participant saw 10 high-sim pairs and 10 low-sim pairs.

Additionally, participants in the same-different condition saw 20 'same' pairs, making 20 'same' and 20 'different' pairs. The training phase used the same 10 pairs of images that were used in Experiment 1.

Procedure The procedure for Experiment 2 was similar to that for Experiment 1. However, images were displayed for a fixed period of 1500ms. After that time the images disappeared. Participants in the same-different condition were presented with a blank screen until they made their decision, while participants in the difference-identification condition where presented with a prompt asking them to type a difference. For participants in the same-different condition, response time was measured from the onset of the presentation of the images. Response time for participants in the difference-identification condition was measured as the time for the first press of a key after the appearance of the prompt asking the participant to type a difference.

Results and discussion

The average response time for each of the two types of experimental image pairs (high-sim and low-sim) was calculated for each participant and the results were analyzed using a repeated-measures Task x Similarity ANOVA. This analysis revealed a reliable interaction between the between-s variable of task (same-different judgment vs. difference-identification) and the within-s variable of similarity (high-sim vs. low-sim). This interaction is shown in Figure 4, $F(1,49) = 4.78$, $MS_e = 3.13$, $p < .05$. As predicted, participants were faster to make a 'different' judgment for low-sim pairs ($M = .78$sec) than for high-sim pairs ($M = 1.04$sec), but were slower to identify a difference between low-sim pairs ($M = 2.58$sec) than between high-sim pairs ($M = 1.99$sec).

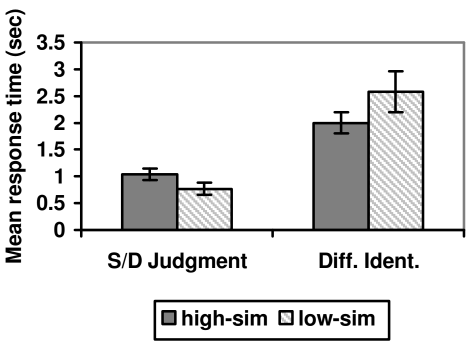

Figure 4: Results from Experiment 1 (error bars represent the standard error of the mean)

Additionally, there were also a reliable main effect for condition ($F[1,49] = 11.04$, $MS_e = 32.79$, $p < .01$). There was no statistically significant main effect for similarity ($F[1,49] = .71$, $MS_e = .46$, *n.s.*). As in Experiment 1, planned contrasts for both tasks revealed that the observed performance differences were statistically reliable in both cases (same-different judgment – $F[1,10] = 19.99$, $MS_e = .377$, $p < .01$; difference-identification –$F[1,39] = 8.52$, $MS_e = 6.95$, $p < .01$).

An item ANOVA provided similar results – a reliable interaction ($F[1,76] = 24.52$, $MS_e = 3.59$, $p < .001$) and a reliable main effects of task ($F[1,76] = 262.98$, $MS_e = 38.46$, $p < .001$) and similarity ($F[1,76] = 4.124$, $MS_e = .60$, $p < .05$).

Experiment 2 replicated the results of Experiment 1. Participants find it easy to identify differences between two highly alignable images, but difficult to decide that these two images differ. Furthermore, this result does not appear to depend on the amount of time participants spend looking at the image pairs, but rather depends on whether a pair of images is alignable or not.

General discussion

Experiments 1 and 2 provide evidence that the participants engaged in different processes when faced with a same-different judgment than when they are concerned with the specific differences that exist between two images. It appears that the similarity of the two objects hinders fast judgments of difference, but aids in the identification of differences between the images.

This dissociation suggests that a same-different judgment can be performed without performing a deep comparison of the compared images – a comparison which would inevitably reveal specific differences between them, if any exist. This is also supported by the fact that participants in the same-different condition were consistently much faster than participants in the difference-identification condition. This difference in the speed of response is consistent with the possibility that a more intricate process is required to identify specific differences than to judge whether two images differ.

The observed pattern of results is consistent with a two-process hypothesis, whose signatures are (1) a rapid same-different judgment process that is fastest (at detecting difference) for highly dissimilar pairs; and (2) a slower difference-identification process that is fastest for highly similar pairs. Pattern (1) is consistent with either a mental-distance model or a feature-intersection model. But these models have no way to predict Pattern (2), nor can they readily explain the disassociation between the two tasks[2].

Pattern (2) is consistent with the predictions and prior findings of structure-mapping theory. Pattern (1) is inconsistent with the predictions that follow from a full structural alignment. However, as noted above, it would follow if we assume that a same-different judgment can be done using a "quick and dirty" readout of approximate overall similarity, as conjectured by Markman and Gentner (2005).

Further Issues

In general, the striking performance differences described in this paper between two seemingly similar tasks suggest that participants utilize a different strategy to perform same-different judgment than to identify specific differences. This poses an interesting question for future research – what is it that makes these two tasks so different?

To address this question, it is helpful to make a further set of distinctions that derive from analogy theory. Rather than a dichotomy between high vs. low similarity, structure-mapping makes a 2x2 distinction: high vs. low structural alignability and high vs. low surface similarity. A pair is structurally alignable to the extent that a structurally consistent interconnected system of relational

[2] One way that these theories might explain Pattern (2), at least in Experiment 1, is by invoking the fact that the low-similar pairs had many differences, while the high-similar pairs had only one. Thus, participants could have been slower on the low-similar pairs because they had to choose between a larger number of differences. However, this explanation will not account for the same pattern in Experiment 2, for which all pairs (high- and low-similar) had a large number of differences.

correspondences can be found between the items. A pair is surface-similar to the extent that the items share most of the surface features, and lack large numbers of distinctive features. A "high similar pair" is one that is *literally similar*: that is, both structurally alignable and surface-similar. A "low-similar pair" is nonalignable (or weakly alignable) and surface-dissimilar.

Studies of the same-different task (including the present one) have traditionally conflated similarity and alignability. By deconflating them, we may be able to discover more about these two processes: for example, perhaps same-different judgments rely on surface similarity, while difference identification relies on alignability. A study by Gentner and Gunn (2001) suggests an approach. They found evidence that listing commonalities lead to elevated performance in a subsequent difference-identification task, while providing a thematic relation diminished performance. Such a method might be used to tease apart the effects of similarity and alignability.

Summary

Our findings suggest that same-different judgments are qualitatively different from the identification of differences. More specifically, the alignability of the images plays a more important role in the identification of differences than in same-different judgments. This dissociation between the tasks is difficult to explain within the scope of mental distance and feature-intersection models.

Because it appears that alignability plays an important role in the identification of differences, it is possible that models of analogy (e.g. Gentner 1983, 2003; Gick & Holyoak, 1983; Holyoak & Thagard, 1989, 1995; Hummel & Holyoak, 1997; Keane, 1988, 1990; Larkey & Love, 2003) might provide interesting insights into the differences that exist between these two intuitively similar tasks.

Acknowledgments

This work was supported by grant N00014-02-1-0078 from the ONR Cognitive Science Program.

References

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1-63.

Farell, B. (1985). Same-Different Judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin, 98,* 419-456.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 195-235). Cambridge, MA: MIT Press.

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory and Cognition, 29*, 565-577.

Gentner, D., & Markman, A. B. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431-467

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5*, 152-158.

Gick, M. L., & Holyoak K. J. (1983) Schema Induction and Analogical Transfer. *Cognitive Psychology, 15*, 1-38

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 29-50.

Harter, J., Ed. (1998) *Plants: 2,400 Copyright-Free Illustrations of Flowers, Trees, Fruits and Vegetables. Dover Pictorial Archive Series.* Mineola, NY: Dover Publications

Holyoak K. J. & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Holyoak K. J. & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Hummel, J. E. & Holyoak K. J. (1997) Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466.

Keane, M. T. (1988) Analogical mechanisms. *Artificial Intelligence Review, 2*, 229-251.

Keane, M. T. (1990) Incremental analogizing: Theory and model. In K. J. Gilhooly, M. T. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought (Vol. 1).* New York: Wiley.

Kurtz K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*, 417-446.

Larkey, L. B. & Love, B. C. (2003). CAB: Connectionist Analogy Builder. *Cognitive Science, 27*, 781-794

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*, 235-249.

Markman, A. B., & Gentner, D. (2005). Nonintentional similarity processing. In R. Hassin, J.A. Bargh and J.S. Uleman (Eds.), *The new unconscious* (pp. 107-137). New York: Oxford University Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Posner, M. I. & Mitchell, R. F. (1967) Chronometric analysis of classification, *Psychological Review, 74(5)*, 392-409.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*, 373-421.

Shoben, E. J. (1983). Applications of multidimensional scaling in cognitive psychology. *Applied Psychological Measurement, 7*, 473-490.

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

Tversky, B. (1969). Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics, 6*, 225-233.