Simulating Time-Course Phenomena in Perceptual Similarity via Incremental Encoding

Andrew Lovett (andrew-lovett@northwestern.edu)
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL 60201 USA

Dedre Gentner (gentner@northwestern.edu)
Department of Psychology, Northwestern University
2029 Sheridan Road, Evanston, IL 60201 USA

Kenneth Forbus (forbus@northwestern.edu)
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL 60201 USA

Abstract

If people are required to respond very quickly in a same-different task, their judgments of sameness are heavily reliant on attribute matches, despite the fact that when given ample time, the judgments seem to rely chiefly on relational matches (Goldstone & Medin, 1994). One interpretation of this temporal pattern is that attribute matches enter into the comparison process before relational matches. However, an alternate explanation, suggested by findings of Sloutsky & Yarlas (submitted) is that attributes are *encoded* before relations. In this case, if the comparison process begins before the encoding is completed, early matches will involve attributes but not relations. We show via a simulation that SME can model the Goldstone & Medin results, as well as the Sloutsky & Yarlas (submitted).

1. Introduction

There is considerable evidence that the processes that govern analogical mapping may also apply to similarity comparisons (Markman & Gentner, 1996; Gentner & Markman, 1997). For example, Markman and Gentner (1996) found that when rating the similarity of two images, subjects attended more to differences connected to the common structure of the two images (*alignable differences*) than to differences unrelated to the common structure. These findings suggest that when asked to find a difference, participants first carried out a structural alignment between the images. Results such as this suggest that the same cognitive process may underlie both analogy and similarity. Consistent with this, the Structure-Mapping Engine (SME) (Falkenhainer, Forbus & Gentner 1989), a computational model of analogy, has successfully modeled perceptual similarity results (Kuehne, Gentner, & Forbus, 2000; Loewenstein & Gentner, 2005).

A critical issue in modeling the psychological processes of analogy and similarity is simulating the time course of processing. In an important study, Goldstone and Medin (1994) found that participants in a similarity task showed relatively greater sensitivity to attribute matches early in processing, and to relational matches later. This suggests that in perceptual similarity computations, attribute matches are made before relational matches. We begin by reviewing this study and then discuss results from Sloutsky and Yarlas (submitted) that suggest that the lag between attributes and relations arises from the time course of encoding rather than

of comparison. We use SME to simulate both studies, providing evidence that supports the time course of encoding interpretation.
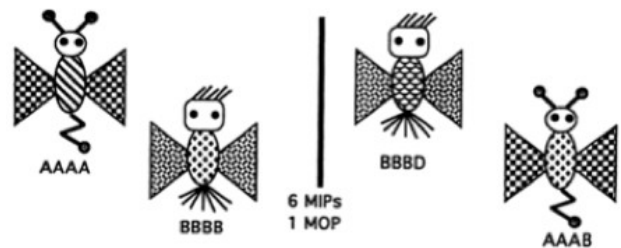
Figure 1. Scenes from Goldstone & Medin (1994)

2. Time-Course Effects in Comparison

In Experiment 1 of Goldstone and Medin's (henceforth G&M) study, participants were told to compare two scenes, each composed of two drawings of butterflies. The butterflies varied along four dimensions: head shape, tail shape, body texture, and wing texture. The two butterflies in the base scene differed on all four dimensions (see Figure 1). The two butterflies in the comparison scene were systematically varied to produce different degrees of feature overlaps with the base two. For example, if the two base butterflies were classified as AAAA and BBBB, where each letter is a value along one of the four dimensions, then a comparison butterfly labeled AAAB would have three features in common with one of the base butterflies and one feature in common with the other. A butterfly labeled BBBD would have three features in common with one of the base scene butterflies and one novel feature.

G&M assessed the similarity between two scenes by looking at subjects' abilities to label them as *different* in a same-different task under a deadline. The assumption was that participants would align each butterfly in the comparison scene with one of the butterflies in the base scene, based on the overall degree of attribute overlap. Participants were told to disregard the butterflies' relative positions; e.g., the top butterfly in one scene could match the bottom butterfly in the other.

There were three deadlines, which varied within-subject: short (1 s), medium (1.84 s), and long (2.68 s). The dependent measure was the error rate on *different* trials: the

number of times subjects mistakenly believed that two different scenes were the same.
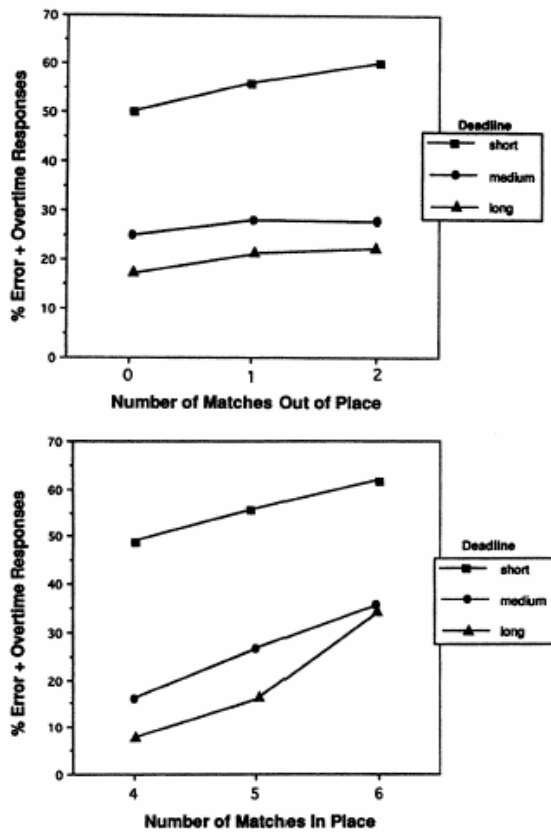

Figure 2. Results from Goldstone & Medin (1994)

G&M analyzed the pairs in terms of matches in place (MIPs) and matches out of place (MOPs). A MIP refers to a match between features in corresponding butterflies. For example, if an AAAB butterfly were matched to the AAAA butterfly, they would share three common features, or MIPs. A MOP refers to a match between features that belong to non-corresponding butterflies: in other words, a cross-mapped attribute. In the example, the final feature of the AAAB butterfly matches that of the BBBB butterfly, producing a MOP. G&M looked at the effect of adding two MIPs with a constant number of MOPs and the effect of adding two MOPs with a constant number of MIPs. At the short deadline, there was no significant difference between increasing the number of MIPs or the number of MOPs (Figure 2). In both cases, as the number of matches increased, subjects were more likely to confuse the two scenes. At the long deadline this relationship changed significantly. The error rate became more sensitive to MIPs and less sensitive to MOPs, with one additional MIP having a greater effect than two additional MOPs.

As Goldstone and Medin noted, these results suggest that early in the process, local attribute matches contributed to a sense of similarity regardless of whether they were structurally consistent with the (yet to be determined) maximal alignment (i.e., whether they were MIPs or MOPs). More generally, these findings could suggest that the comparison process is chiefly sensitive to attribute matches during the early stages of comparison, with sensitivity to relational consistency (e.g., attention to 1-1 correspondences) entering later in the process. Goldstone (1994) simulated these results with his SIAM model. Such an effect might also be captured in SME by assuming that in speeded judgments, an early sense of similarity can be generated from SME's initial parallel matching stage, in which all possible matches are generated between the two items, with no regard for structural consistency (Forbus et al., 1995).

However, before drawing strong conclusions concerning the comparison process, we must consider a second possible interpretation of the G&M findings. The results just discussed could reflect the time course of *encoding*, rather than the time course of *comparison*. Results by Sloutsky and Yarlas (submitted) (hereafter S&Y) suggest that when subjects see an image, their construction of a representation may begin with entities and their attributes, with relations between entities added to the representations later.

We next describe S&Y's findings. Then we describe a simulation of the G&M findings based on the idea that relations may be encoded more slowly than attributes. Finally, we apply this simulation to the S&Y results as well.

3. Evidence for Encoding Effects

In G&M's study, participants had to both encode the two scenes and compare them during the limited time given them. In order to separate encoding from comparison (at least partly), S&Y used a sequential same-different task instead of simultaneous presentation. Subjects first saw the base scene for a limited time, followed by a mask. Then the comparison scene was presented for an unlimited time. Thus, only the time to encode the base scene was limited. Any relational lag must thus be attributed to encoding processes, not to comparison processes.
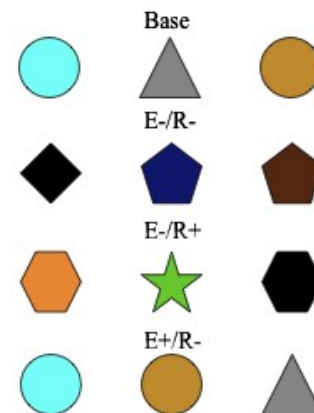

Figure 3. Scenes from Sloutsky & Yarlas (submitted)

The scenes were rows of three objects. All three objects had different colors, but two had the same shape. The shapes appeared in one of three relational patterns: A-B-A, A-A-B, or A-B-B. Figure 3 shows a base scene with an A-B-A pattern. The comparison scene could differ from the base scene either in its elements or in its relations. An *element match* (E+) contained the same three shapes with the same three colors (though not necessarily in the same pattern), while an element mismatch (E-) contained different shapes and colors. A *relational match* (R+) contained the same pattern (e.g., A-B-A / C-D-C), whereas a relation mismatch (R-) contained a different pattern (e.g., A-B-A / C-C-D). S&Y varied the amount of time that the base scene was displayed. In the *ample time* condition, the base scene was shown for 2100 ms, but in the *limited time* condition, it was shown for only 150 ms. The dependent measure was d', accuracy in detecting whether the scenes were different.

The results, shown in Figure 7, indicated that when the attributes were changed (in the E-/R+ and E-/R- conditions), performance was extremely accurate, regardless of whether the relations changed or remained the same. In both cases, performance was only slightly lower with limited encoding time than with ample encoding time. A very different pattern held when the relational pattern changed, but the attributes did not (the E+/R- condition). With ample initial encoding time, performance was high, as in the other two conditions. However, performance dropped sharply with limited encoding time, far more than in the other conditions.

In sum, subjects' accuracy at detecting a change in the relational pattern was high given a long encoding time, and very low given a short encoding time. Their accuracy at detecting changes in attributes was high in both cases. These results suggest that more encoding time was needed for relations than for attributes. Indeed, S&Y conjectured that encoding attributes may be necessary for encoding relations (see also Goldstone, Medin & Gentner, 1991).

Assuming that the process of comparison can begin before the encoding is complete, then matches among attributes may be discovered before the potentially matching relations have all be computed. If so, then an early similarity judgment will be dominated by local attribute matches, without regard for their relational role. As the encoding process continues and relations are added to the representations, then the mapping may be updated using incremental mapping techniques (Forbus et al., 1994; Keane et al., 1988). We explore this possibility in two simulations.

4. Simulation: Incremental Encoding

Our goal in the first simulation was to test whether SME could simulate the G&M results by assuming (a) that attribute encoding precedes relation encoding; and (b) that mapping can begin before the encoding is complete, and be incremented as the encoding proceeds. For the G&M simulation we assumed a two-stage process in which all attributes are encoded first, followed by all relations. This is clearly an oversimplification; in our S&Y simulation we also tested a gradual encoding process. SME can compare

the currently available data at any point in the encoding process, using incremental mapping techniques to update the mapping after new information is encoded.

To simulate performance on the same-different task, we used the *number of differences* produced by SME. This is a reasonable measure because time required to detect that two stimuli are different increases with the number of differences between them (Farell, 1985). Thus given limited time for a comparison, accuracy should increase with the number of differences detected. Interestingly, the accuracy in detecting that two scenes are different does not vary with the number of differences; given a sufficient time to make a decision, subjects' accuracy should remain high regardless of the number of differences.

To measure the number of differences, we used the number of candidate inferences that SME produced. When computing an analogy between base and target cases, SME produces candidate inferences whenever the base contains an expression (connected with the mapping) that is not present in the target. Candidate inferences do not capture non-alignable differences, (i.e., differences not connected at all to the mapping) and so in general they are not adequate for measuring differences. However, they suffice for the simple stimuli used in these experiments. (The stimuli were always completely alignable, so all differences are alignable differences.) Importantly, this measure will also note a difference when, due to time pressure, some information is not encoded in the target.

To avoid hand-coding the stimuli, we sketched the visual scenes for these simulations using sKEA, the *sketching Knowledge Entry Associate* (Forbus & Usher, 2002). sKEA is an open-domain sketch understanding system designed to produce structural representations of a sketch. Objects in the sketch (called *glyphs*) can be identified as instances of categories from a large off-the-shelf knowledge base. sKEA automatically computes various spatial relationships, including relative positions and sizes of glyphs, and has some limited shape recognition capabilities.
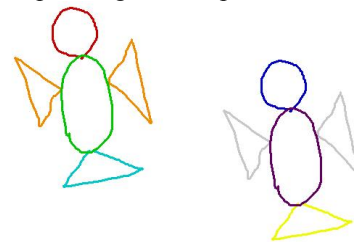

Figure 4. Sketch of G&M stimuli

Simulating Goldstone & Medin's (1994) Results

For our simulation of the G&M study, we drew each of the butterfly body parts (head, wings, body, and tail) as separate glyphs (Figure 4). This captures the fact that the individual parts were perceptually differentiable entities that could match on their own with each other. Thus SME could align parts from corresponding butterflies (MIPs) or non-corresponding butterflies (MOPs).

For convenience, we used color, rather than texture or shape, as the dimension along which all four butterfly parts varied. This is because sKEA can identify colors readily and this choice does not seem to be of theoretical importance. G&M made no distinction between changes in shape, the dimension used for the head and tail, and changes in texture, the dimension used for the body and wings. There were four different colors that could be used for each butterfly part: the color for the first butterfly in the base scene, the color for the second butterfly in the base scene, and two novel colors. To avoid confusion, different sets of colors were used for each butterfly part. This follows a decision made in the original study, in which the texture of the butterfly body never matched the texture of the wings.

To draw the butterflies, we drew glyphs for each part, applying the closest conceptual label from the knowledge base (e.g., the body was labeled `Trunk-BodyCore`). The glyphs for the parts were selected as a group and declared (using sKEA's interface) to be a group glyph, which was given the label `Butterfly`. Most of the visual relationships computed by sKEA were automatically filtered out for this simulation, since subjects were told in the original study that the positions of the butterflies were irrelevant. The glyph group information was used to automatically compute *part-of* relationships between a butterfly and its parts.

For this simulation we sketched one base scene (Figure 4) and 13 comparison scenes, representing the variations of MIPs and MOPs used in the original Goldstone & Medin (1994) study. Because there was no theoretical difference between the different shapes and textures used in the original study and no functional difference between the colors used in our simulation, we were able to use a single base scene without loss of generality. While the original study used short, medium, and long deadlines, only the differences between the short and long deadlines were analyzed in detail, so we used only two deadline conditions. In the short deadline condition, only the attributes of each scene were encoded and fed into SME. In the long deadline condition, an initial SME mapping was built using only the attributes, and then the relations were added and SME remapped. The dependent measure was the number of differences SME found. SME mapped every butterfly part to another butterfly part of the same type and every *part-of* relation to another *part-of* relation, so the only differences found were differences in color. For example, a light blue tail might be matched to a brown tail.

In analyzing our results, we were primarily concerned with the effects of adding two MIPs or MOPs (see Figure 5, and compare to Figure 3). Keep in mind that a decrease in the number of differences identified by SME corresponds to an increase in confusability of the stimuli and thus an increase in error rate in the original study. For robustness, we focus on replicating the ordinal properties of the results in the original studies, as is common in such simulations.

Our results matched G&M's results for human subjects in two important respects. At short intervals, increasing the number of MIPs or the number of MOPs by two had the same influence on similarity. At long intervals, increasing the number of MIPs by two continued to have a strong influence, whereas increasing the number of MOPs by two had a much weaker influence. Thus, we successfully replicated the effect of MOPs increasing similarity more at short intervals than at long intervals. This was the effect of primary interest to us, as it led the original experimenters to conclude that attributes played a stronger role early on in the comparison process than later in the process. We believe this result demonstrates that incremental encoding is a plausible alternative explanation for this effect.
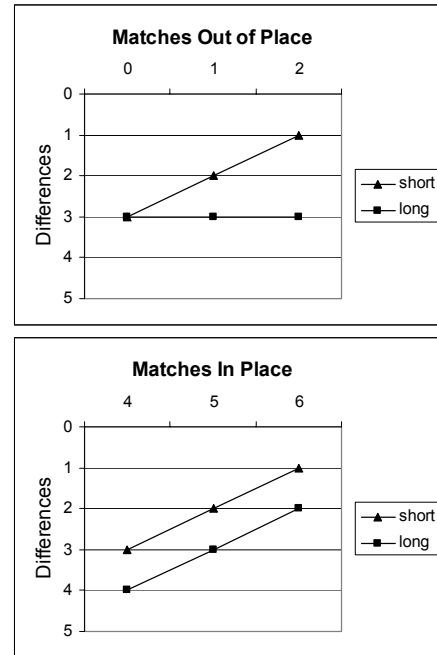

Figure 5. Simulation results for G&M stimuli

We also note that our results differ from the G&M results in a few respects. First, those results showed a greater difference between performance at the short and long intervals. For example, the error rate for 0 MOPs was much higher in the short deadline condition than in the long deadline condition in the original study, whereas in our simulation the number of differences was the same for these two conditions. We suspect that the human results arise from the greater likelihood of decision errors under very short decision deadline—if so, this is a general effect, not specific to the comparison task. Second, G&M found a small effect of the number of MOPs even in the long deadline condition, whereas in our simulation the number of MOPs had no effect at all on the number of differences in the long deadline condition. This result suggests that MOPs may actually affect similarity even when subjects have time to fully encode and compare stimuli. Other studies have also found evidence that MOPs can affect similarity in the absence of time constraints (Goldstone, 1994; Larkey & Markman, 2005). We hope to explore this phenomenon further in a later paper.

Simulating the Sloutsky & Yarlas Results

We tested whether SME could simulate the S&Y results, using the same assumptions of incremental encoding and mapping processes that are launched before encoding has been completed. Given unlimited time to make a comparison, the evidence suggests that even a single difference should generally allow subjects to correctly determine that two scenes are not the same. Thus. our measure of performance for this study was the presence of *any* differences produced by SME.

In S&Y's limited time condition, subjects had limited time to look at the base scene, but had unlimited time to look at the comparison scene while making their comparison. To simulate this, in the limited time condition we limited the number of facts encoded about the base scene, but always encoded every fact about the comparison scene. We entered the base scene as the base case for SME and the comparison scene as the target case, so SME only made inferences from the base scene to the comparison scene. (This ensured that it would find differences when there were facts in the representation of the base scene that were not in the representation of the comparison scene, but not in the other direction, where spurious differences might have been found simply because not all the facts from the base scene had been encoded.)

However, one complexity in the human data should be noted. The prediction from the preceding paragraphs is that subjects should perform at only two levels (mostly correct for E-/R- and E-/R+ (E-) trials and mostly wrong for limited time E+/R- trials). But S&Y's subjects exhibited at least three levels, with medium levels of performance in the limited-time condition for the E- trials (see Figure 7). Of course, in the limited-time trials, it is expected that subjects might fail to encode *some* of the attributes (as well as failing to encode relations). However, on the E- trials, the second scene differed from the first with respect to *all* object attributes, so the only way to explain the lower performance in the limited time condition would be if subjects failed to encode *any* of the attributes. This might be the case. Given that subjects received a large number of trials (60) with no break between trials, they might occasionally have failed to attend during the 150 ms during which the base scene was displayed. We suspect that this accounts for the slight drop in performance from ample to limited time in the E- trials.

To capture these patterns, we varied the number of facts that were encoded for the base scene in the limited time condition. 1/5 of the time, the system failed to encode any attributes. 1/3 of the time, the system encoded all the attributes and a random subset of the positional relations. The rest of the time, the system encoded a random subset of the attributes. We ran each condition 90 times and calculated the percentage of the time that the system found at least one difference between the scenes.

For this simulation, we sketched six base scenes (see Figure 6 for an example) along with the E-/R-, E-/R+, and E+/R- comparison scenes for each base. As in the initial study, E+ indicated a scene in which the three objects shared all their attributes with the base scene, whereas E- indicated a scene with objects possessing entirely different colors and shapes. R+ indicated a scene with objects in the same relational pattern as in the base scene (A-B-A, A-A-B, or A-B-B), whereas R- indicated a scene with objects in a different relational pattern.

We sketched two base scenes for each of the three relational patterns. The R- comparison scenes for these bases each used one of the other patterns. Thus, the six sets of stimuli covered all possible combinations of relational patterns in the base and R- comparison scene. Because there was no theoretical or functional difference between shapes or colors, we were able to use these six stimulus sets for our results without loss of generality. We ran each condition 15 times per stimulus set, producing 90 total trials, and averaged the results.

In contrast to the previous simulation, in this simulation no information about the sketches was entered manually by the user. sKEA automatically determined the color and shape of each glyph, as well as the relative positions of the glyphs, which were encoded as *right-of* relations. Our system also encoded a *same-shape* relation for the two glyphs that shared the same shape.

Our results closely matched S&Y's findings for humans (see Figure 7). As in that study, when there was ample time to encode the base scene, performance in difference detection was roughly equal (and very high) for the E-/R-, E-/R+, and E+/R- pairs. When the base encoding time was limited, our results for the three scenes showed the same divergence as in the human results. For the E-/R- and E-/R+ pairs, there was only a small drop in performance. For the E+/R- pair, the drop was much greater. (We concede that the size of the drop in performance with the E-/R- and E-/R+ pairs was dependent on the probability of encoding attributes from the initial scene, as determined by our probability distribution. However, the ordinal properties of the results were relatively insensitive to changes in that distribution. As long as there is some chance of subjects' failing to encode any attributes and some chance of encoding some attributes but no relations, the results would still replicate the human ordinal results.)

Most importantly, the simulation captures the large advantage of ample time over limited time for pairs with relational differences (E+/R-), and shows that it exceeds the small gain that occurs for ample time over limited time for pairs with attributional differences (E-/R+ or E-/R-). Thus, performance with relations is more sensitive to time constraints than performance with attributes, suggesting that relations are encoded later.

Figure 6. Sketch of Y&S stimuli with A-B-A pattern
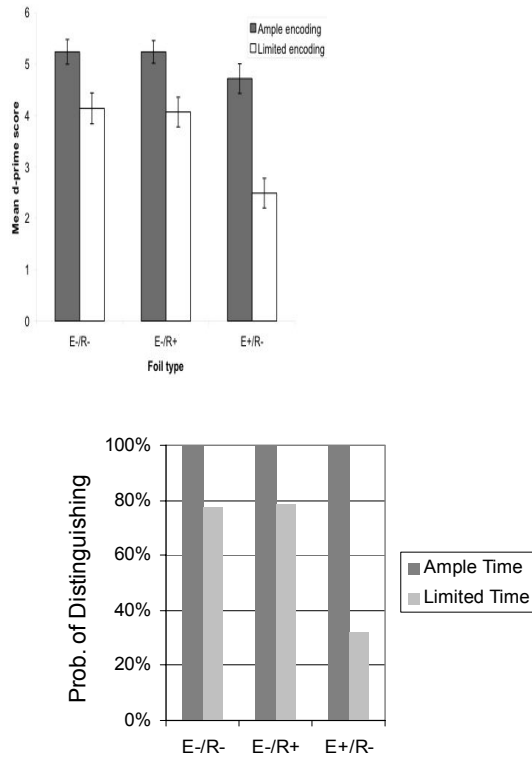
Figure 7. Results of S&Y's study (top) and of our simulation (bottom)

5. Discussion

We believe we have successfully replicated the Goldstone and Medin (1994) and Sloutsky and Yarlas (submitted) studies. Our simulation suggests that the early effects of structurally inconsistent attribute matches (cross-mapped attributes, or MOPs) found by G&M may reflect the time course of encoding, instead of (or in addition to) the time course of comparison itself.

Further research will be necessary to determine the generality of the claim that attributes are encoded before relations. The possibility that the encoding of relations may depend on prior encoding of attributes should also be tested further.

Finally, while our simulations suggest that an incremental comparison process may not be needed to explain the early effect of cross-mapped attributes on similarity, the possibility of such a process remains open. It is possible that attributes have priority *both* during encoding and during comparison. Further studies that independently manipulate encoding time and comparison time are needed to decide this.

References

Falkenhainer, B., Forbus, K. and Gentner, D. (1986). The Structure-Mapping Engine. *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 272-277). San Francisco, CA: Morgan Kaufmann.

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

Forbus, K.D, Ferguson, W. R., & Gentner, D. (1994). Incremental structure mapping. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 313-318). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 42-56.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25*, 524-575.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 3-28.

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

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgment. Cognitive Psychology, 23, 222-264.

Keane, M.T., & Brayshaw, M. (1988). The incremental analogy machine: A computational model of analogy. In D. Sleeman (Ed.), *Third European Working Session on Machine Learning* (pp. 53-62). London: Pitman.

Kuehne, S. E., Forbus, K. D., & Gentner, D. (2004). Modeling infant learning via symbolic structural alignment. *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, (pp. 286-291). Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkey, L. B., & Markman, A. B. (2005). Processes of similarity judgment. *Cognitive Science, 29*, 1061-1076.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50, 315-353.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*(2), 235-249.

Sloutsky, V. M., & Yarlas, A. S. (submitted). Processing of information structure: Mental representations of elements and relations.