

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Learning Relational Categories by Comparison of Paired Examples

Permalink

<https://escholarship.org/uc/item/3943j2wv>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN

1069-7977

Authors

Kurtz, Kenneth J.
Boukrina, Olga

Publication Date

2004

Peer reviewed

Learning Relational Categories by Comparison of Paired Examples

Kenneth J. Kurtz (kkurtz@binghamton.edu)

Department of Psychology, P.O. Box 6000
Binghamton, NY 13902 USA

Olga Boukrina (oboukri1@binghamton.edu)

Department of Psychology, P.O. Box 6000
Binghamton, NY 13902 USA

Abstract

Our central question is whether comparison of co-presented instances promotes category learning. We report results of four experiments testing acquisition of relational categories under conditions of Comparison learning versus traditional Single item learning. In order to control for frequency of exposure, the Single group received twice as many learning trials. Experiment 1 showed more accurate single-item classification at test for both old and new items by the Comparison group relative to the Single group. Experiment 2 used only within-category pairs in the Comparison condition (rather than both types of pairs), but no accuracy advantage was found. Experiment 3 repeated this design using a reduced training set and showed a learning effect of comparison and a marginal advantage in transfer to new items. In Experiment 4, a novel paradigm revealed further evidence of a facilitative effect for within-category comparison. The power of comparison to promote learning and transfer is discussed in terms of mechanisms of encoding and knowledge change.

Introduction

The present research addresses the effect of comparing co-presented instances during classification learning. Nearly all theorists propose that categorizing an instance involves some type of comparison between an instance and stored category representations. A further role for comparison in category learning is between presented instances and remembered instances. Sequential effects may occur if items presented in immediate or near succession are brought into temporal juxtaposition (e.g., Elio & Anderson, 1981). Learners may also experience reminders of previously encountered instances which can guide further processing (Spalding & Ross, 1994; Ross, Perkins & Tenpenny, 1990). Of considerable interest to our project, Ross and Spalding (2000) report that reminding-driven comparisons during category learning mediate attribution of abstract features to individual instances. We investigate the effect of comparison of instances presented together within a classification learning trial with the core prediction of better learning and transfer of relational categories.

This prediction is motivated by several sources including the rich literature supporting the structural alignment account of analogy and similarity (Gentner & Markman, 1997). Perhaps the most directly related evidence is the finding that 4-year-old children extend a label according to

category match more frequently than by perceptual match when the label had been applied to two examples (Gentner & Namy, 1999). After only a single labeled example, children did not favor the category-based extension. Gentner and Namy conclude that a structural alignment process (invited by the common linguistic label) yielded a deeper, more conceptual encoding.

In light of recent findings that classification learning influences similarity, Boroditsky (in press) collected similarity ratings of pairs of object drawings from participants who had first listed either similarities or differences between the items. For both familiar and novel stimuli, items were rated more similar by participants who made comparisons than by those who did not. The effect depended critically on the items being similar – suggesting that comparison drew out a richer realization of the commonalities between alignable objects.

How might comparison work to mediate learning and representation? In the structural alignment framework, comparison influence encoding by: 1) highlighting common relations or alignable differences between examples; 2) projecting candidate inferences from one example to another; 3) promoting abstraction of shared structure as the basis for a generic knowledge structure; and 4) fostering re-representation that alters or re-organizes representational elements in one or both cases (Gentner & Wolff, 2000).

To illustrate, imagine a pair of cases for which a particular relation is encoded in the learner's mental representation of each instance. The process of comparing the representations would render this relation salient and promote abstraction and transfer (e.g., Loewenstein, Thompson & Gentner, 1999). Now, consider a common relation that is differently encoded in each case. Re-representation is posited as a means of aligning non-identical relational structures when there is semantic overlap (Gentner & Kurtz, in preparation) or a computational opportunity (Yan, Forbus, & Gentner, 2003). Next, consider a relation that has only been encoded in the representation of one of two instances. If there is sufficient surrounding structure in common, then a candidate inference would be projected from the more fully elaborated case to the sparser one (Gentner, 1983; Markman, 1997).

Finally, consider a common relation that is not encoded in the mental representations of either case. The mechanisms

listed above depend on the presence of relational information encoded in item representations. In the current project, we study learning in a novel domain and the underlying relation defining each category is far from self-evident to the uninitiated. In fact, short of resorting to exemplar memorization, the learning task is best characterized as trying to discover each relation. We posit a role for comparison in the discovery of relational content.

The notion of *manifest* versus *latent* representational content is of use here (Clement, Mawby, & Giles, 1994). While an individual may have somewhere in their idea of ‘dog’ the knowledge that dogs often feast on foodstuffs fallen to the floor during a family meal, this relational content is probably not routinely activated in a context-independent manner (Barsalou, 1982). Therefore, an analogy between a dog and vacuum cleaner might initially fall flat for someone who does not have the ‘cleaning-up-of-table-scrap’ aspect of their ‘dog’ concept activated. However, a thorough comparison of dog and vacuum cleaner could well activate latent matching content. Such *resurfacing* occupies a place between novel inference and highlighting, but, like the others, it relies on available relational content.

What is needed is a mechanism for articulating relational content over presumably unstructured initial inputs. Such processing is likely be of critical importance in any type of routine formation of structured mental representations since constraints are needed on which of the vast range of possible relations among objects, scenes, and situations in everyday experience should be explicitly encoded. As a number of theorists have put forth, language may be of particular use with regard to this problem.

We propose that comparison provides potential for a kind of *side-to-side* (as contrasted with top-down or bottom-up) interpretation process that promotes relational construal. The best evidence we can draw upon is the phenomenon of analogical bootstrapping in which intensive comparison of two partially understood depictions of a simple physics principle (heat flow) led participants to a deeper, more relationally-rich construal (Kurtz, Miao, & Gentner, 2001).

It is not clear how such analogical insight occurs, but here are two speculations. The first is consistent with the notion of progressive alignment (Kotovsky & Gentner, 1996) and states that observed commonalities at the level of lower-order representational elements (i.e., attributes, objects, first-order relations) may serve as entry points from which a familiar higher-order relation can be invoked. An example of instantiating a richer representation would be going from: *is-high(square)* and *is-low(circle)* to: *is-above (square, circle)*. The second speculation is that relations do not need to be built up so much as they need to be picked out of a crowd. The idea here is that many relations hold for any given case; too many to routinely articulate and encode. When given an opportunity to compare cases, the potential arises to find a manageable intersection of the relations.

In order to explore the power of comparison in knowledge change, our experimental question is as follows: can

comparison promote the acquisition of novel categories defined by non-obvious relations? Relational categories have been treated theoretically (Gentner & Kurtz, in press; Markman & Stilwell, 2001) and have begun to receive empirical attention (Kurtz & Gentner, 2001; Rehder & Ross, 2001).

In a study using an early version of the present paradigm, Kurtz & Gentner (1998) found that participants reached a learning criteria for classification accuracy more quickly with trials consisting of within-category pairs than with single-instance trials. However, this can be attributed either to comparison or to more frequent exposure to training items, i.e., two instances per trial versus one. This creates a difficult circumstance for the researcher since fully convincing evidence for a comparison effect in learning (with frequency of exposure controlled) requires obtaining reliably higher classification accuracy on the basis of half the number of trials. This is the challenge we pursue.

An additional purpose of this project is the advancement of greater naturalism in the study of categorization. The dominant paradigm is a two-way classification task with instances that are clearly dimensionalized sets of perceptual or verbal features. Our stimuli are line drawings depicting a set of realistically varying “rock arrangements” having no clear reduction into a compositional set of underlying dimension values. Learners are asked to acquire three different categories to avoid two limitations inherent in binary classification: 1) a perfect success rate can be achieved based on an ability to identify examples of only one category and; 2) task demands encourage hypothesis-testing for a boundary over positively-defined concepts.

Experiment 1

One major concern in designing the first study was ensuring that the Comparison condition actually elicited comparison. An act of comparison can be shallow or intensive, and this difference can be a causal factor (Kurtz, et al., 2001). A failure to observe a comparison effect might be due to a failure by participants to compare. Classifying a within-category pair can easily be done with consideration of only one of the instances. Therefore, instead of all same-category pairs, we designed the Comparison condition to use an equal mix of within- and between-category pairs. In this mixed-pairs version, the status of any given pair is not known to the learner. Since the two instances may or may not belong to the same category, we collect two separate classification judgments on each learning trial. Accordingly, the participant must give direct consideration to each member of the pair. It is implicit in the task that the participant must consider whether or not to guess the same category for the two instances in each trial. However, classifying each instance could still be done largely independently despite taking place in a common task space.

For this reason, we used an orienting task at the beginning of each learning trial to encourage comparison. Participants were asked to consider the role played by one of the rocks relative to the rest of the arrangement and then to look for a

corresponding rock in the other instance. The orienting task for Single learners was to consider the role of one of the rocks in the arrangement. In both conditions this unenforced orienting task (no response was collected) was followed by a question that did require a response: whether or not the participant found the orienting task helpful. This was to discourage participants from ignoring the orienting task.

Method

Participants A total of 100 undergraduate students at Binghamton University received a course credit.

Materials A set of 36 images of rock arrangements was created on the computer. Rocks in each arrangement varied in color, shape, and size. A subset of 24 images were designated as the training instances and the remaining 12 images constituted the transfer set. The rock arrangements were evenly distributed across three categories given the names: “Tolar,” “Besod,” and “Makif.” The category Tolar was defined by the presence of two stacked rocks similar in color and shape. Besod was defined by the presence of one rock supported by two others. Makif was defined by monotonically decreasing height from left to right. Care was taken that each instance conformed to exactly one of the relational categories. For the Comparison condition, a fixed set of pairings was established with an equal number of within-category and between-category pairs.

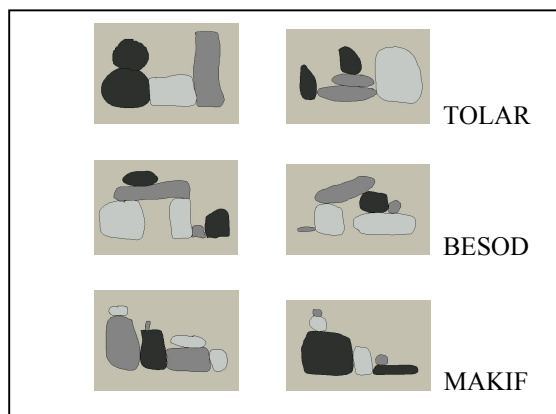


Figure 1: Sample Rock Arrangement Stimuli used in Experiments 1-3 Shown in Same-Category Pairs

Procedure Each participant was randomly assigned to one of the two conditions. Before the learning phase, participants read a set of instructions including a cover story about different rock arrangements created by the “Ladua” culture. Ss were instructed to try to learn to tell which rock arrangements belonged to which of the three types.

In the Single condition ($n=50$) an attempt was made to minimize the potential for temporal comparison by using a pseudo-random order in which each trial showed an instance from a different category than that of the previous trial. On each of the 48 learning trials participants were shown a

single instance from the training set on the computer screen along with the orienting task: “Study the example, then focus on a single rock and consider the role it plays in the arrangement.” Participants gave a forced-choice response regarding the helpfulness of the orienting task and were then asked to classify the rock arrangement into one of the three categories. After their choice was entered, corrective feedback was provided for a fixed interval of 3s. The stimulus image remained on screen for the entire trial.

In the learning phase of the Comparison condition ($n=50$), instances were presented two-at-a-time for 24 trials. On each trial, the presentation of the two instances on either the left or right side of the screen was randomized. The orienting task instructions for each trial were: “Study the examples, then focus on a single rock in one of the examples and consider the role it plays in that arrangement. Try to decide which rock plays a corresponding role in the other example.” Participants made helpfulness judgments as in the Single condition. Ss were then asked to classify one of the instances followed by the other. Whether the left or right instance was queried first was alternated by trial. After the second response, corrective feedback for each of the responses was presented simultaneously for a total of 6s.

The learning phase in both conditions was followed by a common testing phase. Participants were presented with 24 old and 12 new items in random order and asked to classify each in a single-instance trial without feedback. Additional dependent measures were subsequently collected, but space limitations prevent their inclusion in this report.

Results and Discussion

The learning data reveal that it was not easy for most participants to acquire the relational categories in the allotted number of trials. We note that a set of pilot data showed that performance did not increase notably with twice the training. It is, however, important to remember that chance is 33.3% percent on a three-way classification, so the accuracy data reflects considerably more learning than it would appear at first glance. First we describe the learning data though we did not conduct statistical tests since the critical comparison between conditions is performance in the test phase when all participants respond to the same type of trial (single instance). Early (first quarter) classification accuracy shows that Comparison learners ($M = .44$, $SD = .18$) got off to a slow start compared to the Single group ($M = .52$, $SD = .19$), but they caught up by the final quarter: Single ($M = .70$, $SD = .22$) and Comparison ($M = .69$, $SD = .25$).

In the test phase, Comparison ($M = .75$, $SD = .22$) was significantly better than Single ($M = .65$, $SD = .23$) on old instances, $F(1, 98) = 4.73$, $MSe = .234$, $p < .05$. In addition, a transfer effect was found with Comparison ($M = .72$, $SD = .22$) significantly more accurate than Single ($M = .59$, $SD = .27$) in classification of novel instances, $F(1, 98) = 7.29$, $MSe = .444$, $p < .05$. In sum, while Comparison learning presented a similar level of challenge during acquisition, a

reliable comparison advantage was found at test compared to Single learners receiving equal exposure.

Experiment 2

The goal of the second study was to determine whether a comparison effect would be found using only within-category pairs and a single categorization response per trial. The structural alignment view predicts a greater likelihood of comparison-driven effects on learning and encoding given the opportunity to compare alignable examples. However, as discussed, it is difficult to pin down the extent to which participants invoke a comparison process when making a joint classification response.

Method

Participants A total of 95 undergraduate students at Binghamton University received a course credit.

Materials The materials were the same as in Experiment 1. The assignment of pairs for the Comparison condition was accomplished by random generation of within-category pairings for each participant.

Procedure Each participant was randomly assigned to one of the two conditions. The Single condition ($n=46$) was conducted as in Experiment 1. The Comparison condition ($n=49$) followed the procedure of Experiment 1 except that participants were trained only on within-category pairs. Unlike Experiment 1, participants made a joint classification choice in response to both of the instances on each trial. Corrective feedback for the one response was shown for 3s.

Results and Discussion

A much different result was obtained relative to the findings of Experiment 1. Comparison learners with only within-category pairs showed good performance in the learning phase ($M = .63$, $SD = .16$) as compared to Single learners after an equal number of trials ($M = .53$, $SD = .14$), but not after an equal number of exposures ($M = .60$, $SD = .14$). No significant differences were found in test performance on old items (Comparison: $M = .66$, $SD = .23$ and Single: $M = .67$, $SD = .21$; $p > .8$) or transfer to new items (Comparison: $M = .61$, $SD = .23$ and Single: $M = .63$, $SD = .24$; $p > .6$). The evidence suggests that mixed pairs offer a more productive learning context than exclusively within-category pairs. This could be a benefit derived from evaluating whether or not co-presented pairs are from the same category. It could be due to useful contrastive evaluation of different-category items. However, it is worth noting on this point that no reliable difference was observed between learning accuracy on within-category and different-category trials in the Comparison condition of Exp. 1 ($p > .3$). Therefore, we are inclined to consider additional explanations. One possibility is that the joint classification task failed to fully encourage comparative evaluation of both instances in the trials. A final and somewhat compelling possible culprit is the 3s window for evaluating

feedback as opposed to the 6s window for the dual-feedback in the Comparison condition of Exp. 1. In the current study, Comparison learners actually had half the overall amount of time to study images with their correct labels then was provided to Single learners.

Experiment 3

Given the lack of comparison advantage in Experiment 2, we considered the question of which is better: many different within-category comparisons over a large training set or repeated within-category comparisons over a small training set? It has been shown that larger category size promotes better transfer to novel examples when the instances in the training set are sufficiently variable (Homa & Vosburgh, 1976). Our hypothesis was that in the case of relational categories, repeated comparisons of within-category pairs in a smaller set would actually be more likely to promote an advantage of comparison in transfer accuracy. If Comparison learners in Exp. 2 underachieved due to a failure to fully compare and/or insufficient feedback time, repeated training on fewer examples might prove more conducive to comparison-driven learning.

Method

Participants A total of 87 undergraduate students at Binghamton University received a course credit.

Materials The materials were the same as in Experiment 1 except that the total number of examples in the training set was reduced to 12. Each category was represented by four, rather than eight, instances. All possible within-category pairings appeared once (determining 18 of the 24 trials in the Comparison condition). The remaining 6 trials were randomly determined for each participant including exactly one exposure of each training item.

Procedure The procedure was the same as in Experiment 2, though the same number of learning trials with fewer items in the training set yielded more exposures to each instance in the Single ($n=42$) and Comparison ($n=45$) conditions.

Results and Discussion

Comparison learners ($M = .79$, $SD = .16$) showed excellent overall accuracy on learning trials with the small category size relative to Single learners ($M = .68$, $SD = .16$). Although we have emphasized test performance rather than learning accuracy, Comparison learners were reliably more accurate across learning trials, $F(1, 85) = 10.39$, $MSe = .25$, $p < .005$ with equal frequency of exposures. The Comparison group ($M = .68$, $SD = .16$) also performed better on transfer items, $F(1, 85) = 3.91$, $MSe = .18$, $p = .051$, though the significance here was marginal. In performance on old items at test, a trend was found ($F(1, 85) = 2.54$, $MSe = .09$, $p = .11$) favoring the Comparison condition ($M = .83$, $SD = .16$) over Single ($M = .76$, $SD = .21$). These results provide yet another turnaround—the previous failure to find an advantage of comparison using within-category pairs is

overturned in the case of repeated comparisons with a small set of training items. The advantage is not limited to overlearning of the items in the small set since the results at test including transfer to new items favor Comparison. Our interpretation is that repeated comparison opportunities among increasingly familiar instances better allows the fruits of comparison to be borne out.

Experiment 4

We developed an additional paradigm to evaluate comparison of instances in category learning. The key difference is that in all conditions the orienting task is dropped and each learning trial begins with a single-instance classification judgment. In the Single condition, feedback is provided and the trial is done. In the Comparison condition, a within-category context item appears next to the target item, and the participant is asked for a second time to classify the initially presented target. Once the learner has made their second response, feedback is then provided based on the final response. Therefore, the only difference between conditions is that learners in the Comparison group are asked to repeat their classification choice in light of the availability of a context item from the same category for their consideration. We believe this is a naturally motivating and “unforced” version of comparison. A further advantage of this design is that since only the single target item is classified in each condition, we are better able to evaluate the impact of comparison during the learning phase.

Method

Participants A total of 50 undergraduate students at Binghamton University received a course credit.

Materials A full-sized stimulus set was used as in Experiments 1-2, but some alterations were made to the set. It was decided that the Tolar category was of a somewhat different character than the other categories since only two rocks in the entire arrangement participated in the relation of “same shape and color of two stacked rocks.” In the other categories, the relation was more globally realized in the overall arrangement. A new relational definition and item set for the Tolar category was developed in terms of a symmetrical outline for each arrangement across the vertical axis. In addition, instances of the Makif and Besod categories were fine-tuned to ensure that the relation was globally realized in each rock arrangement. For example, the relation “one rock supported by two” would not be localized in a set of small rocks off to the side. These modifications were expected to make the learning task somewhat easier and the results more interpretable. Pairings for the Comparison condition were assigned such that all 28 possible same-category pairs in each category occurred once during learning and the remaining 12 learning trials were repetitions equated for instance exposure.

Procedure Each participant was randomly assigned to one of three conditions. Prior to the start of the learning phase participants read the instruction set. However, in order to

help limit cases in which a learner embarked on a counterproductive approach, the instruction set was given the following addition: “Each of the three types is based on a distinct way of arranging rocks. Please note: It is not a small detail or a feature of one single rock. It is something about the way in which the group of rocks are arranged.”

In the Single condition ($n=20$), participants completed 96 classification learning trials of single instances of the new set of rock arrangements in pseudo-random order. Feedback was given after each trial with study time self-paced rather than a fixed window. In the Comparison condition ($n=17$), each trial began exactly like a Single condition trial. However, participants did not receive feedback on their response. Instead they were shown another within-category instance from the training set as a context item. Participants were asked for a second time to classify the initial target item (this was reinforced by presenting the question under the target, not the context item). An accompanying instruction encouraged Ss to compare the target to the additional example from the same category. Participants were instructed to feel free to change their initial answers or not. Ss received feedback on their second response with self-paced study time.

A third condition called Identical ($n=13$) was conducted just as the Comparison condition except that the additional context item was a repeat of the target item—resulting in two identical images shown side-by-side. This task (responding twice to the same stimulus) was justified in the instructions as something Ss might find helpful.

In all conditions, participants went on to a test phase like that used in Experiments 1-3.

Results and Discussion

Means and standard deviations for the learning phase accuracy are shown in Table 1. One-way ANOVA showed a main effect of learning condition on classification accuracy in the first quarter, ($F(2, 47)= 5.22, MSe = .16, p < .05$), in the last quarter, ($F(2, 47)=3.19, MSe = .12, p < .05$), and in the overall performance, ($F(2, 47)= 3.77, MSe = .11, p < .05$). The first quarter difference was driven by the Identical group and most likely reflects participants adjusting to the somewhat odd repeated query. Planned comparisons showed that last-quarter accuracy (the final 24 trials) was significantly higher in Comparison versus Single, $t(35)= 2.22, p < .05$, as well as Comparison versus Identical, $t(28)= 2.27, p < .05$.

In the test phase there was a marginal main effect of learning condition on accuracy, $F(2, 47)= 3.167, MSe = .142, p = .051$. Planned comparisons showed that Comparison learners performed significantly better on old items ($M= .93, SD= .15$) than Single learners ($M = .82, SD = .18$), $t(35)= 2.06, p < .05$. Performance on new items was also better for the Comparison condition ($M = .87, SD = .18$) than the Single condition ($M = .70, SD = .20$), $t(35)= 2.71, p < .05$. Trends (presumably due to small sample size) were found in favor of Comparison over Identical for old items ($p = .08$) and new items ($p = .12$). No difference was found in

accuracy between the Single condition and Identical condition. We see in these results good evidence for better learning with the opportunity to compare to a within-category context item versus conditions with no additional comparison or a kind of item self-comparison that serves as a full control for exposure (equal number of classification responses; equal number and duration of item exposures).

Table 1. Classification Accuracy in Learning.

	Mean	SD
First quarter		
Single	.69	.18
Comparison	.71	.17
Identical	.52	.16
Last quarter		
Single	.81	.18
Comparison	.94	.15
Identical	.77	.24
Overall		
Single	.77	.16
Comparison	.84	.16
Identical	.67	.18

General Discussion

We conclude that comparison of instances during category learning is not necessarily of great impact, but when task constraints emerge that engage the learner to apply the machinery of comparison, superior performance in learning relational categories is achieved. These findings are most naturally understood in terms of learning to construct richer, more sophisticated encodings of category instances. While this is a difficult process, it is made easier by comparison.

Acknowledgements

We wish to thank Dedre Gentner for her important role in the work leading up to this research. We thank the members of the Learning and Representation in Cognition (LaRC) Laboratory including Aliza Nelson for constructing stimuli.

References

Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10(1), 82-93.

Boroditsky, L. (in press). Comparison and the development of knowledge. *Cognition*.

Clement, C., Mawby, R., & Giles, D. (1994). The effects of manifest relational similarity on analog retrieval. *Journal of Memory & Language*, 33(3), 396-420.

Elio, R. & Anderson, J.R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 397-417.

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

Gentner, D. & Kurtz, K.J. (in press). Learning and using relation categories. In Ahn, W.K., Goldstone, R.L., Love, B.C., Markman, A.B., & Wolff, P.W. *Categorization inside and outside the lab*. Washington, DC: American Psychological Association.

Gentner, D. & Kurtz, K.J. (in prep). Relational and object matches in on-line evaluation of sentence analogies.

Gentner, D. & Markman, A.B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45-56.

Gentner, D. & Namy, L. (2000). Comparison in the development of categories. *Cognitive Development*, 14, 487-513.

Gentner, D. & Wolff, P. (2000). Metaphor and knowledge change. In E. Dietrich, & A. Markman, *Cognitive Dynamics: Conceptual Change in Humans and Machines* (pp. 295-342). NJ: Lawrence Erlbaum Associates.

Homa, D. & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning & Memory*, 2(3), 322-330.

Ross, B., Perkins, S., & Tenpenny, P. (1990). Reminding-based category learning. *Cognitive Psychology*, 22, 460-492.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797-2822.

Kurtz, K.J. & Gentner, D. (1998). Category learning and comparison in the evolution of similarity structure. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 1236.

Kurtz, K.J. & Gentner, D. (2001). Kinds of kinds: Sources of category coherence. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 522-527.

Kurtz, K. J., Miao, C., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, 10(4), 417-446.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, 6, 586-597.

Markman, A.B. (1997). Constraints on analogical inference. *Cognitive Science*, 21, 373-418.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Intelligence*, 13, 329-358.

Rehder, B. & Ross, B. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(5), 1261-1275.

Spalding, T. & Ross, B. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20(6), 1251-1263.

Spalding, T. & Ross, B. (2000). Concept learning and feature interpretation. *Memory & Cognition*, 28, 439-451.

Yan, J., Forbus, K., Gentner, D. (2003). A theory of rerepresentation in analogical matching. *Proceedings of Twenty-Fifth Annual Meeting of Cognitive Science Society*.