Title
Comparisons in Category Learning: How Best to Compare For What

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Higgins, Erin
Ross, Brian

Publication Date
2011

Peer reviewed

Comparisons in Category Learning: How Best to Compare for What

Erin Jones Higgins (eljones3@illinois.edu)
Department of Psychology, 603 E. Daniel St.
Champaign, IL, 61820 USA

Brian H. Ross (bhross@illinois.edu)
Department of Psychology, 603 E. Daniel St.
Champaign, IL, 61820 USA

Abstract

Learners frequently compare across examples of categories in order to learn more about them. Surprisingly little work has been done looking at how the type of comparison one engages in during learning affects what information is eventually learned. Across three experiments we show that the type of comparison, either within-category or between-category, affects what conceptual information is acquired and the benefits of each type of comparison change depending on the type of category one is learning.

Keywords: comparisons, category learning, problem solving

Introduction

Category knowledge is used in a variety of ways and there is evidence that the way categories are learned affects how they are represented (Markman & Ross, 2003). Specifically, the particulars of the learning task are critical for determining what information a learner acquires (e.g. Jones & Ross, in press). This view of the learner as actively processing items suggests the potential importance of other kinds of active processing. Comparison of examples has often been suggested as important for category learning (e.g. Spalding & Ross, 1994), but the various possibilities for comparisons have not been adequately explored. A better understanding of learning requires a better understanding of comparison effects. In this paper, we suggest that the type of comparison affects later performance, but its effect is modulated by type of category.

Most comparisons research examines how comparisons enable better analogical transfer. In these experiments the focus is on how comparing across examples benefits analogical reasoning contrasted with not comparing at all, rather than focusing on different types of comparisons. When learners are told to compare across examples, successful transfer to novel problems is higher than when comparisons are not part of learning (e.g. Gentner, Lowenstein, & Thompson, 2003).

It is not the case that any comparison will benefit learning. Rittle-Johnson and Star (2009) demonstrated that the type of information being compared when solving math problems affects what is learned. Learners who compared different solution methods to the same problem showed greater conceptual knowledge and flexibility than learners who compared different problem types.

These results suggest that the type of comparison matters for learning tasks that rely on comparisons. It is clear that comparisons are critical to category learning, so the type of comparison could also be important to consider. It is possible that between- and within-category comparisons provide the learner with very different information about the category or categories being learned, and one or the other could be beneficial depending on the goals of the task and what information the learner needs to take away from the comparison. This observation has yet to be addressed in the category learning literature.

Gentner's (1983) structure-mapping theory proposes that people reason through analogies by noticing common relational structures of the two instances being compared (abstracted from the surface features of the examples). In addition, it is easier to notice differences between two examples when they are highly similar compared to when they are not similar because highly similar examples share many alignable differences (Markman & Wisniewski, 1997; Gentner & Markman, 1997). How this theory plays out in category learning is not yet clear. On the one hand, structure-mapping theory might predict that within-category comparisons are more beneficial than between-category comparisons because items from the same category will be more similar to each other and have more alignable differences. On the other hand, if the categories being compared have many alignable differences, then between-category comparison could also be helpful for learning the feature values for each category. Additionally, there are categories whose relational structure is not difficult to acquire. For these types of categories, between-category comparisons may also be more helpful if learning to distinguish between categories is critical.

There is some evidence suggesting that between-category comparison is more beneficial for category learning. Kornell and Bjork (2008) had participants learn categories of paintings. When examples from one category were spaced between examples from other categories, learners were better able to classify novel items from those categories at test compared with novel items from categories whose instances were massed together during learning. The examples were presented one at a time, so participants were never told to make comparisons, however, the benefit for spaced items suggests that they did benefit from having examples from multiple categories in mind.

There are reasons that within- and between-category comparisons could be helpful, though it seems that the impact of each comparison type could be modulated by various learning conditions. It is critical to gain a better understanding of when each comparison type is helpful and why. One possibility is that the type of category that needs to be learned could determine the benefit of one comparison type over another. We begin with one explanation regarding why category type would matter, but we acknowledge that there may well be other explanations that make similar specific predictions.

If the relational structure of the items is difficult to learn or requires abstraction away from the surface properties of an example, then it is critical to focus on what each category is like before attempting to discern the properties that distinguish one category from another. Comparing within-category would be more effective for allowing the learner to notice the common relational structure of category members. On the other hand, if the category structure is closely tied to the surface or perceptual features of the category then it may be easy enough to learn the relational structure regardless of comparison type. Consequently the learner will benefit from between-category more than within-category comparisons, as between-category comparisons highlight how one category differs from others.

There are two goal of the present study. One, to demonstrate that the type of comparison (within-category or between-category) affects what category knowledge is acquired. Two, to see if the type of category being learned determines which comparison type is most beneficial.

In the first experiment, we test the idea that categories whose structures are closely tied to their perceptual features will benefit more from between-category comparison than within-category comparison. Finding that they do, we attempt to replicate the finding in a second experiment using real world materials and a different learning task. In the third experiment, we examine our other claim, that categories whose structure requires abstraction away from the surface properties of an example will benefit more from within-category comparison, by teaching learners two different categories of math problems through either within-category or between-category comparison.

Experiment 1

In Experiment 1 we investigated whether categories whose structure is tied to the perceptual properties of the exemplars would be better learned through between- or within-category comparison. Learners were taught categories that are commonly used in category learning research—two categories of artificial aliens, Deegers and Koozles. Participants either learned the categories by comparing two exemplars from the same category or one exemplar from each category. After learning, participants performed a classification test that included old and new items.

Method

Design Participants were randomly assigned to one of two between-subjects conditions: between-category comparison learning or within-category comparison learning. Two different combinations of features were used and were counterbalanced across participants.

Participants Participants were 43 undergraduates from the University of Illinois who participated for course credit. Three participants did not follow instructions, so only 40 participants are considered in the analyses below.

Materials Stimuli for both the learning and test phases were two types of fictitious aliens, called Deegers and Koozles, which varied along six binary features: arms, tail, antennae, legs, eyes, and mouth (See Figure 1). For each feature, there was a prototypical Deeger value and a prototypical Koozle value. The structure of the categories was family resemblance. Items shown during learning always had five out of six features in common with the category prototype and one feature in common with the other category's prototype to reflect variation among members. Four of the features were diagnostic of category membership, meaning one value appeared 75% of the time in items in one category and only 25% of the time in items in the other category, while the two remaining features were consistently the same value across both categories' items. This resulted in 4 items per category. The diagnosticity of features was counterbalanced across participants, as noted above.

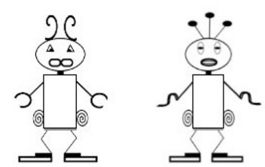

Figure 1: Example alien prototype stimuli for one subject

On each trial, two items were compared. There were 12 unique comparisons for the within-category condition and 16 unique comparisons for the between-category condition.

The two features that did not vary during learning (legs and tails in Figure 1 prototypes) were varied at test to generate additional exemplars. The prototypes were also shown. This resulted in 32 exemplars (16 per category) at test. Of the 32 exemplars, 8 were old and 24 were new.

Procedure After giving informed consent, participants were told that they would be learning about two different categories of aliens, Deegers and Koozles, and would later be tested on what they had learned.

On each trial during the learning phase, participants were shown two items with their category labels. At the top of the screen, the between-category comparison learners saw the prompt: "List how this Deeger [Koozle] and this Koozle [Deeger] are the same and different." The within-category comparison learners saw the prompt: "List how these Deegers [Koozles] are the same and different." To make the category labels salient, "Koozle" was always presented in red and "Deeger" was always presented in blue. Participants typed their response and then clicked on a box to submit the response and move to the next trial. There were 24 trials. Pairs of items were presented in a random order for each participant.

After the learning phase, participants were given instructions about the test phase. On each trial, they saw one alien and two category labels. Their job was to choose the correct category label for the alien using the mouse. After making a selection, the next trial appeared 500 ms later. There were 32 trials. Items were presented in a random order for each participant. Finally, participants were debriefed and thanked for their time.

Results & Discussion

As predicted, learners who engaged in between-category comparison during learning had higher overall classification accuracy (M = 0.70, SD = 0.16) than learners who engaged in within-category comparison (M = 0.60, SD = 0.16), t (38) = 2.16, $p < 0.05$. The same results were also obtained for only old items (M = 0.69, SD = 0.18 for between-category comparison learners; M = 0.58, SD = 0.18 for within-category comparison learners), t (38) = 2.05, $p < 0.05$. There was a marginal effect of comparison type for only new items (M = 0.71, SD = 0.17 for between-category comparison learners; M = 0.60, SD = 0.17 for within-category comparison learners), t (38) = 1.94, p = 0.06.

For categories whose structure is salient in the perceptual features of its members, between-category comparison is more beneficial for learning to classify category members. These results are consistent with Kornell and Bjork's (2008) finding that categories whose members are spaced between other category's members during learning are learned better than categories whose members are massed together.

Experiment 2

It was critical to replicate the results of Experiment 1 using a different set of materials. If we want to make the claim that between-category comparison is more beneficial for categories whose structure is salient in the perceptual features of its members then we need to be able to demonstrate the effect across a wide range of perceptual stimuli. Here we chose to use a real world set of categories whose perceptual properties are critical to category membership—birds. One critical difference between the bird categories and the artificial stimuli used in Experiment 1 is that the birds have more feature variation and complexity, making hypothesis testing more difficult.

In Experiment 1, participants were explicitly told to compare across items. When learning in the real world, a learner is not always prompted to compare examples. One question that arises is whether comparison, and more specifically the type of comparison, matters when it is incidental to the task and no specific analysis of similarities and differences is required. To address these issues, we had participants do a modified same/different judgment task while learning about real-world categories of birds. Participants were presented with a bird picture and its label and then asked to choose the label for another bird picture. The two label choices were one that was the same as the first bird's label and one that was different. Comparison was not required for this task, though it was extremely helpful.

Finally, we wanted to see if comparison only occurs if items are presented at the same time. To test this, we manipulated whether items appeared simultaneously or sequentially.

Participants learned about six different categories of birds: three in a way that encouraged between-category comparison and three in a way that encouraged within-category comparison. They either learned about the bird categories by viewing two exemplars simultaneously or sequentially. To determine whether one comparison type was more beneficial, as well as whether comparison benefits only arise when items are side-by-side, all participants performed a novel classification test after learning. Novel classification is a typical means of assessing category knowledge, as it addresses the degree to which participants are able to generalize what they had learned[1].

Method

Design Participants were randomly assigned to one of the two between-subjects conditions: either all bird exemplars were presented on the screen in pairs (Pairs condition) or they were all presented one at a time (Singles condition).

In addition, there was a within-subject manipulation across all participants: out of six bird categories participants had to learn, three were learned in a way that encouraged between-category comparison and three were learned in a way that encouraged within-category comparison.

Participants Participants were 48 undergraduates from the University of Illinois who participated for course credit.

Materials Stimuli for the learning and test phases were color images of birds from six families in the Passeriformes order, compiled by Jacoby, Wahlheim, and Coane (2010) from www.whatbird.com. The six bird families were finches, flycatchers, swallows, thrushes, vireos, and warblers (See Figure 2). For each family, 12 exemplars were used in the experiment: six were used in the study phase and six were used for the novel classification test.

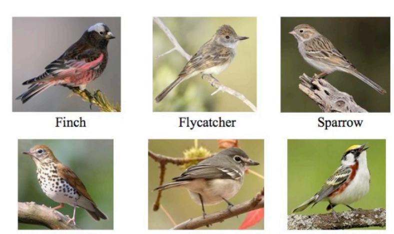

Figure 2: Example bird stimuli

Procedure After giving informed consent, participants were told that they would be learning about six different

[1] We would like to thank Larry Jacoby and Chris Wahlheim for their helpful suggestions for this experiment.

categories of birds and would later be tested on what they had learned.

On each trial during the learning phase, participants in the pairs condition were presented with two birds. Below the bird on the left side of the screen was its label (e.g. "This is a Finch") and below the bird on the right side was a prompt (e.g. "Is this a Finch or a Warbler?"). The participant's job was to determine which of two labels went with the bird picture on the right side of the screen and click on that label. One option was always the same label as the left-side bird's label. Essentially participants were making a same/different judgment, only they chose either the same or a different category label instead. Participants were never told that they should use the other bird or compare across instances.

For categories of birds in the within-category comparison condition, the correct answer to the prompt was always the same label as the label for the bird on the left. For categories of birds in the between-category comparison condition, the correct answer to the prompt was always the label that did not match the label for the bird on the left. The list orders were created such that the correct answer was not predictable by the response that preceded it.

After making a response, the participant received feedback in the form of "Correct!" or "Incorrect". Regardless of the correctness of the participant's response, the two bird pictures along with their correct labels and the corrective feedback remained on the screen for 12 seconds.

The only major difference for the singles condition was that the bird pictures were presented one at a time. First, participants viewed a bird and its label (e.g. "This is a Finch") for eight seconds. Next, participants viewed a different bird and a prompt (e.g. "Is this a Finch or a Warbler?"). The participant had to determine which of two labels went with the bird, and one option was always the same label as the preceding bird's label. Feedback was similar to feedback in the pairs condition, except that only one bird was shown with its label and the feedback remained on the screen for six seconds. As in the pairs condition, participants were never told that they should use the other bird or compare across instances.

There were 36 learning trials. For between-category comparison bird categories, an exemplar from each category appeared equally often with each of the other between-category comparison categories.

After learning, participants made family level category-learning judgments for novel exemplars. Participants used a number pad and indicated on a scale of 16% (chance) to 100% how well they had learned that family of birds. Next, they participated in a novel classification test. On each trial, participants were given an exemplar and asked to choose one of the six family labels presented on the screen using the mouse. After choosing a label, participants indicated how confident they were in their selection on a scale between 16% (chance) and 100%. (No differences were found between conditions for confidence ratings, so they are not discussed further.) There were 36 trials, presented in a different random order for each participant. Finally, participants were debriefed and thanked for their time.

Results & Discussion

Learning Performance Learning performance was calculated as the proportion of correct responses across all learning trials, with chance being 0.05. 2x2 mixed factorial ANOVA showed an interaction between presentation type and comparison type, F (1, 46) = 4.13, $p < 0.05$, so follow up paired t-tests were run. For singles learners, between-category comparison trials were marginally more accurate (M = 0.73, SD = 0.14) than within-category comparison trials (M = 0.67, SD = 0.10), t (23) = 1.73, p = 0.10. For pairs learners, there was no difference between comparison types, t (23) = 1.00, $p > 0.05$.

Overall learning accuracy was marginally higher for pairs learners (M = 0.75, SD = 0.10) than singles learners (M = 0.70, SD = 0.08), F (1, 46) = 3.51, p = 0.07. There was no main effect of comparison type on learning accuracy (between-category items: M = 0.73, SD = 0.13 and within-category items: M = 0.71, SD = 0.12), F (1, 46) = 0.758, $p > 0.05$. The main learning finding is that when items are presented simultaneously, learning accuracy is marginally better than when they are presented sequentially (singles); however, comparison type did not affect learning accuracy across either presentation type.

Test Performance The main goal of this study was to see if between-category comparison was more beneficial for category learning even when it was incidental to the task. In addition we wanted to know if the method of presentation—pairs or singles—affected the degree to which participants benefited from comparisons at test. Test performance was calculated as the proportion of correct responses across trials, with chance being 0.167 (participants had to choose between 6 possible category labels). A mixed factorial ANOVA showed a main effect of comparison type. Classification accuracy for novel items in categories that were learned through between-category comparisons (M = 0.42, SD = 0.18) was higher than items in categories learned through within-category comparisons (M = 0.36, SD = 0.19), F (1, 46) = 4.88, $p < 0.05$. There was a marginal main effect of presentation type. Pairs learners were marginally more accurate (M = 0.43, SD = 0.16) than singles learners (M = 0.35, SD = 0.16), F (1, 46) = 2.83, p = 0.10. There was no interaction between presentation type and comparison type, F (1, 46) = 0.70, $p > 0.05$, suggesting that learners benefit from between-category comparisons regardless of whether they view the items simultaneously or sequentially (See Table 1).

Table 1: Classification accuracy (and s.d.) broken down by presentation type and comparison type

	Between-category	Within-category
Pairs condition	0.46 (0.19)	0.39 (0.17)
Singles condition	0.37 (0.16)	0.32 (0.21)

In a task where comparisons were incidental to the learning task and specific analyses of similarities and differences across items were not required, the type of comparison was still critical for learning. Bird families that were learned in a situation that encouraged between-category comparison were learned better than families learned in a situation that encouraged within-category comparison. These results replicate the findings from Experiment 1 using a very different learning task and materials set, suggesting that between-category comparisons are beneficial for categories whose structure is salient in the perceptual features of their members.

Experiment 3

We have now demonstrated that the type of comparison matters when learning about categories. The second goal of our study was to determine if different types of categories benefit from different types of comparisons. There are reasons to think that within-category comparisons are more helpful when the structure of the category is abstract and difficult to learn. For these categories, learners must go beyond the surface or perceptual features of the category members to acquire the relational structure. According to structural alignment, alignable differences between instances are helpful for seeing the relational structure that is common across items. If there are more alignable differences between items within the same category, then within-category comparisons should be more beneficial. To test this explanation, we turned to an abstract problem-solving domain—mathematics.

In this experiment we taught learners about two different mathematical concepts, permutations and combinations. Participants either compared within-category (two permutations problems) or between-category (a permutations problem and a combinations problem). At test we had learners classify novel problems by determining which formula was appropriate. For other novel problems, we gave the formula and the learners had to apply it and solve the problem. We predicted that within-category comparisons would be more beneficial than between-category comparisons, as evidenced by higher classification and problem solving accuracy.

Method

Design Participants were randomly assigned to one of the two between-subjects conditions: within-category comparison learning or between-category comparison learning.

Participants Participants were 12 undergraduates from the University of Illinois who participated for course credit.

Materials Four problems, two permutation and two combination, were written for the learning phase (see Ross, 1984, Appendix 2 for examples of similar problems). Within each problem type, one problem referred to animate objects while the other referred to inanimate objects. Regardless of the between-subjects manipulation, an animate object problem was always paired with an inanimate object problem for a comparison.

Four problems, two of each type, were written for the classification test. Another four, two of each type, were written for the problem-solving test. An equal number of animate and inanimate object problems were written for each type and each test.

Procedure After giving informed consent, participants were told that they would be learning about two different categories of math problems and would later be tested on what they had learned. They were given a packet that included 7 pages and a cover sheet. On the first page after the cover sheet was a sheet that detailed how to solve a particular class of problems (e.g. permutations). On the second page was a problem of that type. Participants had to identify the variables and apply the formula, but did not have to solve the arithmetic. Once satisfied, they could turn the page and see the solution. This process repeated either for a problem of the same type, if the participant was in the within-category comparison condition, or for a problem of a different type (e.g. a combinations problem) if the participant was in the between-category comparison condition. After completing the second problem, the last page had the following prompt: "List the similarities and differences between the problems and their solutions." Participants could write as much or as little as they wanted and take as long as they needed to get through the entire packet. When finished, the participant turned in the packet and received a second packet, which was in the same format. Across the two packets there were a total of four problems, which resulted in two comparisons (one per packet).

After learning, participants were given another packet that included both tests. The first test was the classification test. Participants were shown four novel problems one at a time and had to circle which of the two formulas (permutations or combinations) was the correct one. The formulas were appropriately marked with a P or a C, so even if the participant had not memorized the formulas, he or she should have been able to recognize which was which. The second test was the problem-solving test. Participants were shown four novel problems one at a time along with the appropriate formula. Their task was to solve the problem with the exception of doing the arithmetic. They could take as much time as they needed. Finally, participants were debriefed and thanked for their time.

Results & Discussion

Learners who were given within-category comparisons during learning had higher classification accuracy at test ($M = 0.83$, $SD = 0.13$) than learners who were given between-category comparisons ($M = 0.54$, $SD = 0.25$), $t(10) = 2.57$, $p < 0.05$. This finding suggests that when learning about categories requires learners to abstract away from the surface features of the examples, within-category comparisons are more beneficial for acquiring conceptual knowledge.

Within-category comparison learners showed better, (though not reliably better) problem solving ability (M = 0.91, SD = 0.13) than between-category comparison learners (M = 0.75, SD = 0.27), t (10) = 1.35, $p > 0.05$. Although the difference was not statistically significant, this may be due to ceiling performance by some participants in both groups.

The findings from this experiment suggest that within-category comparisons are beneficial for learning about categories that require abstraction away from the surface features of the items. In some recent work not reported here, we find similar results using very different materials that also require abstraction away from the surface features of the items (see Erickson, Chin-Parker, & Ross, 2005 for example materials).

General Discussion

Across three experiments we have shown that the type of comparison used during learning affects what is learned about categories. We showed this across a variety of different categories as well as different learning tasks. Additionally, the benefits of each type of comparison are modulated by the type of category being learned.

If the relational structure of the items is not evident in the items' surface features, then comparing within-category leads to better learning. This explanation fits within the structural alignment view (Gentner, 1983), which argues that examples sharing more alignable differences will be more effectively compared. If two examples are from different categories and share fewer alignable differences then within-category comparisons will be more effective.

On the other hand, when the structure of the category is based on the perceptual or surface level features of its members, as in the bird categories and artificial alien categories used in the present experiments, then it is relatively easy to determine without needing to compare within-category. For these categories, between-category comparisons will be more beneficial, assuming the goal is to learn what distinguishes different categories, because the distinguishing information is more salient when comparing across multiple categories.

This is a first step toward understanding when and why certain types of comparisons are more beneficial than others. We proposed one explanation here, but we acknowledge that there are other factors that could affect how comparison type influences learning. For instance, the degree to which the categories being learned are similar to each other may play a role in determining which comparison type is useful. If two categories are similar then they will share many alignable differences potentially making between-category comparisons effective for learning.

In addition, when learning about categories, the goal is not always just to classify. Future work should address how comparison type affects other measures of learning, for instance inference ability and explanation generation. Additionally, more work needs to be done to determine other factors beyond category type that determine which kinds of comparisons are beneficial.

References

Erickson, J.E., Chin-Parker, S., & Ross, B.H. (2005). Inference and classification learning of abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 86-99.

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

Gentner, D., Lowenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393-408.

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

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Retrieval practice enhances natural category learning: Effects on accuracy and on metacognition at the levels of items and categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *36*, 1441–1451.

Jones, E.L. & Ross, B.H. (in press). Classification versus inference contrasted with real world categories. *Memory & Cognition.*

Kornell, N. & Bjork, R.A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585-592.

Markman, A.B. & Ross, B.H. (2003). Category use and category learning. Psychological Bulletin, 129, 592-613.

Markman, A.B. & Wisniewski, E.J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 54-70.

Rittle-Johnson, B. & Star, J.R. (2009). Compared to what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology,101*, 529-544.

Ross, B.H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16,* 371-416.

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