Task Influences on Category Learning

Huichun Zhu (huichun.zhu@gmail.com)
Department of Philosophy, Carnegie Mellon University
Pittsburgh, PA 15213 USA

David Danks (ddanks@cmu.edu)
Department of Philosophy, Carnegie Mellon University
Pittsburgh, PA 15213 USA; and
Institute for Human & Machine Cognition
Pensacola, FL 32502 USA

Abstract

Two plausible influences on category learning are presentation format and the learner's beliefs about future category uses. Standard experimental designs typically do not manipulate these two dimensions independently, and so their effects cannot be easily disentangled, particularly since both plausibly affect task difficulty. In the present paper, we independently manipulated the two influences, and found that they have different effects on learning difficulty and learned category representation.

Keywords: category learning; presentation format; belief effects.

Introduction and Related Research

Concepts are widely recognized as crucial for cognition, whether they are the "glue that holds our mental world together" (Murphy, 2002, p. 1) or the "building blocks of thought" (Solomon, Medin, & Lynch, 1999, p. 99). The two standard experimental formats for category learning are classification and inference learning. In the classification learning format, people are shown all of the feature values for a case, and then asked to infer the category. They might, e.g., be shown a picture and asked whether it is a cat or a dog. In contrast, the inference learning format requires people to infer the value for some feature given the category label and values of the other features. They might be told that some animal with an obscured head is a dog, and then asked about the shape of its ears. Anderson (1990, 1991) conjectured on theoretical grounds that the learning format would (or should) not influence category representations. Recent empirical work suggests that learning format matters (e.g., Anderson, Ross, & Chin-Parker, 2002; Chin-Parker & Ross, 2002, 2004; Markman & Ross, 2003; Ross, 1996, 1997, 1999; Yamauchi & Markman, 1998). This paper aims to separate the influence of learning format and goal beliefs.

This separation is complicated by the fact that prior work has found differences in learning difficulty between classification and inference formats (e.g., Yamauchi & Markman, 1998). There are at least three possible sources of these differences. First, there may be intrinsic task difficulty differences between classification and inference learning formats (as argued in Yamauchi & Markman, 1998). That is, it may actually be harder to learn (when all else is equal) using one learning format rather than another. When testing Anderson's conjecture, we should not try to control such variations in task difficulty due to learning format.

Second, the differences in learning difficulty could be due to differences in the underlying statistical structure for the predictions. The target of learning varies between the two formats: in classification learning, one is learning to predict the category; in inference learning, the various features of the object provide the learning targets. Thus, even if the overall statistical structure of the category (i.e., the joint probability distribution over the class and features) is the same for both learning formats, we have no guarantees that the statistical structures for the two *learning problems* are the same. For example, Yamauchi & Markman (1998) use the same underlying category structure for both learning format conditions. However, in the inference learning format, the relationship between any target (i.e., feature) and the single category label is deterministic, while in the classification learning format, the relationship between the target (i.e., category) and any single feature is probabilistic. Presumably, task difficulty and category learning can be affected by prediction certainty, which is experimentally controllable but is unrelated to Anderson's conjecture.

Third, the learning target in the inference learning format almost always varies from case to case: the feature that participants must predict in one case is not necessarily the feature that they must predict for the next case in the learning sequence. Thus, if a category has more than one feature, classification learning will allow for significantly more repetition of particular cases than inference learning. Consider a typical category learning experiment with one binary category label and n binary features. (Throughout this paper, we will denote a particular case by $LF_1F_2 \ldots F_n$, where L is the category label and F_i the i^{th} feature, and use '?' to denote the variable to be predicted in a particular case.) In the classification learning format, each learning exemplar has a fixed presentation: $?F_1 \ldots F_n$. In the inference learning format, however, each exemplar can be presented in n different ways: $L?F_2 \ldots F_n$, …, $LF_1F_2 \ldots ?$. The inference learning format thus has fewer repetitions of particular presentations, and so the learning problem is potentially harder. The potential difficulty here is due not to intrinsic features of inference learning, but rather to the fact that participants are asked to predict different features for different learning cases. Furthermore, classification learning participants arguably need only to learn a single conditional probability distribution: namely, the probability of the category label given the features (i.e., $P(L|F_1F_{2...}F_n)$. In contrast, inference learning participants must track either the full joint distribution, or n distinct conditional distributions: $P(F_1|LF_{2...}F_n)$, …, $P(F_n|LF_{1...}F_{n-1})$. Like the second source of task difficulty difference, variation in category learning caused by mental load and repetition is not evidence against Anderson's conjecture.

Both of these latter two influences on task difficulty arise from between-format differences in the learning target, but they are separable. For example, Nilsson and Olsson (2005) used an experimental design in which both classification learning and inference learning were probabilistic, thereby mitigating the second potential influence. Their results were significantly different from prior work, which suggests that variation in the statistical structure of the *learning* problem is an actual influence on category representation. At the same time, they failed to remove the asymmetry in mental load and repetition, since they varied the feature to be predicted across cases in the inference learning format.

In order to balance task difficulty for the classification and inference learning formats and to properly address Anderson's conjecture, we argue that one should use what we call a *Fixed Inference Format* in which participants are asked to infer the value of a feature given the class and other feature values, but where the target feature does not change from case to case. This learning format clearly does not suffer from the third potential source of task difficulty. In addition, it is much simpler to balance the prediction certainty problem for Fixed Inference format and traditional classification learning, since one has a fixed learning target in this format.

In addition, prior work has not carefully separated out the influences of presentation format, and of participants' beliefs about their overall goals and future category use. Exposure to one particular stimulus structure or presentation format during the learning phase presumably leads participants to form beliefs about the future tasks with which they will be faced. That is, the presentation format generates beliefs about the test phase goal, and those beliefs can potentially affect category representations. A more complete understanding of the role of learning format thus suggests that one should determine whether beliefs about subsequent category use (i.e., about the learning goal) directly influence the learned category representations.

Experiment

The overall goal of the present experiment is to separate out the influences on learned category representations of (i) presentation format and (ii) beliefs about subsequent use. The experiment is (in some ways) exploratory; in particular, participant behavior given Fixed Inference Format cases is currently unknown. The category structure used for the learning phase of the experiment is shown in Table 1.

An important characteristic of this non-linear category structure is the statistical symmetry between L and F_1. Specifically, $P(L|F_1) = P(F_1|L)$ and $P(L|F_1F_2F_3F_4) = P(F_1|LF_2F_3F_4)$. Note that any pair of categories with statistical symmetry between L and F_1 and a probabilistic relationship between the category label and any feature will not be linearly separable.

Table 1: Category structure for learning phase

Category 1 Exemplars	*1*111	Category 0 Exemplars	*1*110
	*1*101		*0*100
	*1*100		*0*011
	*1*011		*0*010
	*0*110		*0*000

The experimental design is between-participant with 3 × 2 conditions: three learning formats crossed with two goal beliefs. The learning formats are: Classification format (CF), in which the prediction target is always the category label; Fixed Inference format (FIF), in which the prediction target is always feature F_1 (the italicized feature in Table 1); and Random Inference format (RIF), in which any feature can be the prediction target during learning (i.e., the target can vary from case to case). Although we believe that FIF is the more appropriate contrast for CF (as argued above), we include the RIF conditions as a replication control.

The CF vs. FIF design allows us to address Anderson's conjecture properly. If a label formally equals a feature, and different presentation formats do not affect learning as long as the statistics are equal (as argued in Anderson 1990, 1991), then the CF and FIF conditions should not have different test phase performance. On the other hand, differential test phase performance suggests that a category label has different informational content than a feature; in other words, presentation format directly affects learning (and not just through differential task difficulty).

We used two goal belief conditions: Classification goal (CG), in which participants are explicitly instructed that they will have to predict the category label for some novel instances after learning; and Inference goal (IG), in which participants are explicitly instructed that they will have to predict feature values for some novel instances in the test. Participants from all conditions are told that only the final goal is important and (when applicable) they are encouraged to ignore the fact that the actual format of learning might differ from the focus of the goal. The different goals were described in the experiment cover story, and reminders of the goal were displayed with each learning case.

Participants

122 Carnegie Mellon University students participated as part of a series of unrelated experiments. They were compensated $10 for the series, which took 45 to 60 minutes to complete. The experiment was carried out on computers in the Laboratory for Symbolic and Educational Computing at Carnegie Mellon.

Design and Materials

The experiment focused on classification of (imaginary) insects. The cover story asked participants to play the role of biologist and learn to distinguish between two kinds of bugs. The bugs were differentiated by the values of four binary features (eyes, legs, wings and tails). There are ten learning exemplars (see Table 1), and the learning phase consisted of a series of blocks, where each block contained each exemplar once in random order. The experiment used supervised learning in all conditions: see a case, predict a value, receive feedback, study the fully displayed case, and then move to the next case. The learning phase had at least 4 blocks (i.e., 40 cases) and finished when the participant either obtained 90% accuracy within a block, or finished 30 blocks (300 cases). All phases of the experiment were self-paced. Note that there are six possible combinations of feature values that were never shown in the learning phase. Test phase judgments about those cases provide additional information about the learned category representation.

In the test phase, participants were given both inference and classification tests. For the eight inference tests, participants were asked the eight possible questions of the form: "Given that this bug belongs to category [*0/1*], what value for its [F_i] is most likely?" (In the questions provided to participants, all variables and values were replaced with their actual names.) Participants responded using a slider that ranged from 100 (F_i definitely attains value 1) to 0 (F_i definitely attains value 0). For the sixteen classification tests, participants were presented a bug with all features and asked questions in the form of "How likely is it that this bug is a member of category [*1*]?" Participants responded using a slider that ranged from 100 (definitely a member of category [*1*]) to 0 (cannot be a member of category [*1*]). The first test phase block was always the one that was consistent with the stated goal for that participant. In other words, if the participants were in a classification goal condition, then they first saw the classification task in the test phase; in the inference goal conditions, they first saw the inference task in the test phase.

Results and Discussion

Because we aimed to separate out the influence of several different factors, we report our results in three sections.

Task Difficulty. For conditions with the same format and different goals, there were no significant differences in the number of cases required to reach 90% performance (two-sample t-tests). Since goal does not seem to affect task difficulty, we pooled the conditions to examine the effects of presentation formats on difficulty level. Figure 1 shows the percentages of participants reaching the 90% correctness threshold (rather than the 300-case limit) for the three learning formats: 90%, 70% and 50% for CF, FIF and RIF respectively. There is a significant difference between pooled Classification format and pooled Random Inference format (p = .021; binomial test), but not between pooled Classification format and Fixed Inference format (p = .149), or between Fixed Inference format and Random Inference format (p = .344). Because participants who failed to reach 90% accuracy in any block presumably failed to learn the category structure, we exclude them from the remaining data analyses.

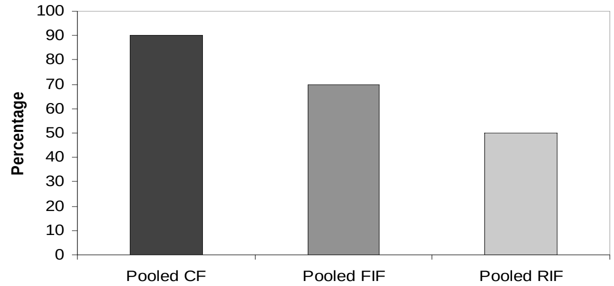

Figure 1: Percentage of participants reaching 90% accuracy in the learning phase

We ran one-way ANOVA on the mean of the number of cases required to reach 90% performance. The means of the three pooled conditions are 95.56, 120 and 154.29 (Figure 2). There is a significant difference among the three pooled conditions (p = .009; One-way ANOVA). In particular, Tukey HSD post hoc test shows that there is a significant difference between Classification format and Random Inference format (p = .006). The other differences—Classification format vs. Fixed Inference format, and Fixed Inference format vs. Random Inference format—are non-significant (p = .332, p = .194 respectively).

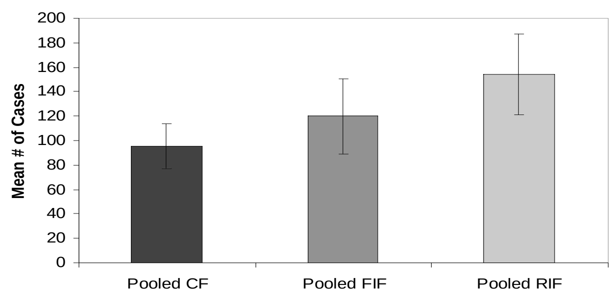

Figure 2: Mean number of cases to reach 90% accuracy

Both difficulty analyses conclude that it is significantly easier to learn in Classification format than in Random Inference format. This replicates the findings of earlier research that compared the difficulty of these two formats (e.g., Chin-Parker & Ross, 2002; Yamauchi & Markman, 1998). However, since the Fixed Inference learning format is not significantly more difficult than the Classification learning format under either analysis, the difference between Random Inference and Classification formats does not seem to be due simply to learning format differences. At least part of the difference is likely due to variation on other dimensions (e.g., the increased memory burden in the Random Inference learning format).

Format and Goal Main Effects. For six of the ten learning exemplars, the <Classification format, Classification goal> condition had the most accurate mean test phase classification rating, which is significantly more than chance (p = .002, binomial test). At a slightly coarser level of analysis, a condition with the Classification format yielded the most accurate mean classification ratings for seven of the ten training exemplars, which is significantly more than chance (p = .019, binomial test). Similarly, conditions with the Classification goal yielded more accurate mean test phase classification ratings for nine of the ten training exemplars, which is significantly (p = .021, binomial test) more than chance.

At a finer grain of analysis, we found a main effect (2-way ANOVA) of the Presentation Format on mean test phase classification ratings for three learning exemplars (Figure 3): *0000* (p = .001); *1110* (p = .037); and *0110* (p = .017). (Since there were no interaction effects for these exemplars, we compared pooled ratings to determine learning format effects.) In particular, Tukey HSD post hoc test showed that, for exemplar *0000*, there is a significant difference between Classification and Fixed Inference formats (p = .001), and between Classification and Random Inference formats (p = .003) with Classification format as more accurate in both pairs. For exemplar *1110*, there is a significant difference between Fixed Inference and Random Inference formats (p = .011), with Fixed Inference format as more accurate. Finally, for exemplar *0110*, there is a significant difference between Classification and Random Inference formats (p = .005) with Classification format as more accurate, and a slightly significant difference between Fixed Inference and Random Inference formats (p = .065), with Fixed Inference format as more accurate. All other pairwise differences were not significant.

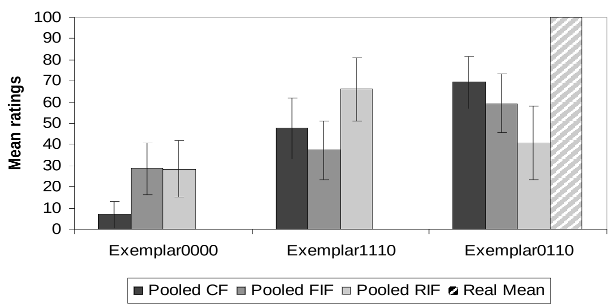

Figure 3: Learning exemplars with format effects in classification test phase

We found a main effect (2-way ANOVA) of Goal on two learning exemplars (Figure 4): *0011* (p = .029) and *0100* (p = .047). For both, Classification goal participants were more accurate in mean test phase classification ratings.

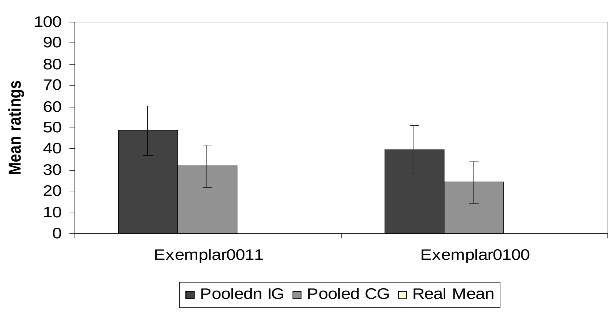

Figure 4: Learning exemplars with goal effects in classification test phase

We also ran a two-way ANOVA to examine whether Goal or Presentation Format leads people to classify the six transfer exemplars differently. We found a main effect (2-way ANOVA) of the Presentation Format factor on the transfer exemplar *1001* (p = .003) (Figure 5). Tukey HSD post hoc test showed that Classification format is significantly different from both Fixed Inference (p = .004) and Random Inference (p = .044), with Classification format participants reporting higher likelihoods that *1001* belongs to category *1*. A mild main effect (2-way ANOVA) of the Presentation Format factor was also found for the transfer exemplar *0001* (p = .092) (Figure 5). Tukey HSD post hoc test showed that the rating of Classification format is mildly different from that of Random Inference format (p = .075), with Classification format participants giving a higher likelihood that *0001* belongs to category *0*.

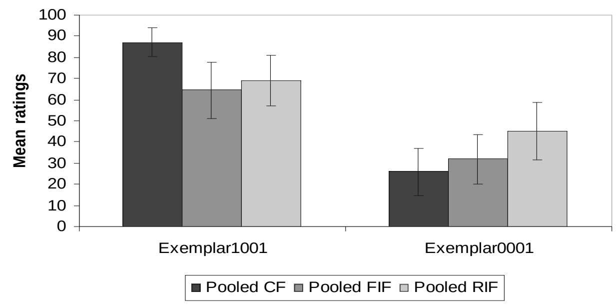

Figure 5: Transfer exemplars with format effects in classification test phase

Overall, the main effect analyses of the test phase performance indicate that both Presentation Format and Goal belief influence how people learn categories, though the effects are not necessarily strong or omnipresent. Classification format leads to better performance in classification tasks when compared to Fixed Inference format and Random Inference format. Classification goal also seems to lead to better performance in classification tasks when compared to Inference goals.

Format and Goal Interaction and Consistency Effects. We found an interaction effect of Goal and Presentation Format on learning exemplar *0010* (p = .029) (Figure 6). Pair-wise comparisons showed that responses in the <Classification format, Classification goal> condition are significantly different from those in the conditions with (i) Fixed Inference format and Classification goal (p = .007); and (ii) Random Inference format and Classification goal (p = .042). For both, <Classification format, Classification goal> yielded better performance.

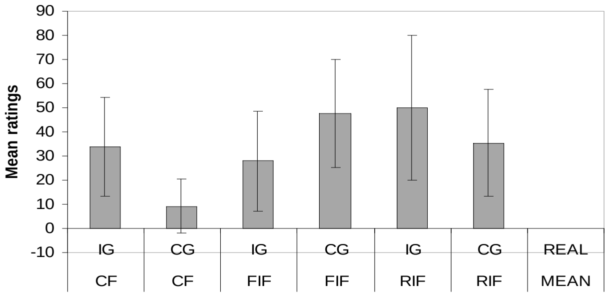

Figure 6: Interaction effect for learning exemplar *0010* in classification test phase

For the transfer exemplars, we found an interaction (2-way ANOVA) of the Goal factor and the Presentation Format factor on the transfer exemplar *0101* (p = .048). Tukey HSD post hoc test showed that participants in the <Classification format, Classification goal> condition gave a significantly lower mean likelihood that exemplar *0101* belongs to category *1* than did those in the <Random Inference format, Classification goal> condition (p = .039) (Figure 7).

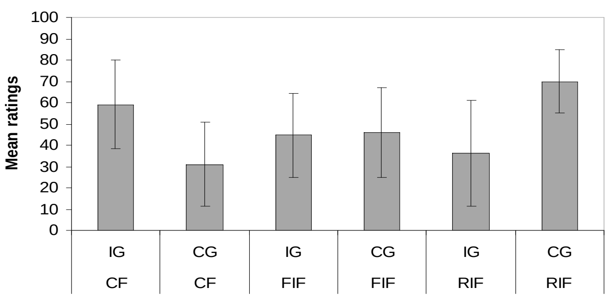

Figure 7: Interaction effect for transfer exemplar *0101* in classification test phase

For the inference test phase, two-way ANOVA revealed an interaction effect (p = .042) of Goal and Presentation Format on ratings of the likelihood of F_1 in individuals from category *1* (Figure 8).

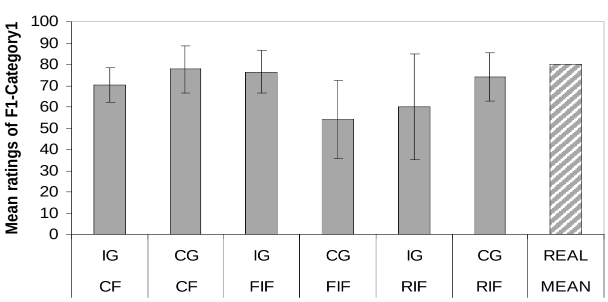

Figure 8: Interaction effect in inference test phase for feature F_1 given category *1*

Perhaps the most interesting interaction effect arises in the Fixed Inference format conditions. In those conditions, participants consistently learned to predict the same feature, (i.e., F_1: the Wings) for all learning exemplars. Presumably, by repeatedly focusing on F_1 throughout the learning phase, participants in Fixed Inference format should have learned F_1 considerably better than participants in other conditions. Figure 9 reveals a different pattern. Of all six conditions, <Fixed Inference format, Inference goal> leads to the second-best performance for inferring F_1 for category 1. In contrast, <Fixed Inference format, Classification goal> leads to the worst performance. The same pattern was repeated for test phase ratings of F_1 for category 0. (Note that separate ratings were obtained for F_1 for the two different categories.) Again, <Fixed Inference format, Inference goal> leads to the second-best performance, while <Fixed Inference format, Classification goal> leads to the worst performance. Given that these two conditions have the same presentation format (i.e., Fixed Inference format), the difference in the inference task should plausibly be attributed to the Goal factor. Within Fixed Inference format, possessing the Inference goal leads to significantly better performance on this inference task than possessing the Classification goal. This is naturally understood as a consistency effect: learning is impaired when the beliefs about final goal conflict with the presentation format.

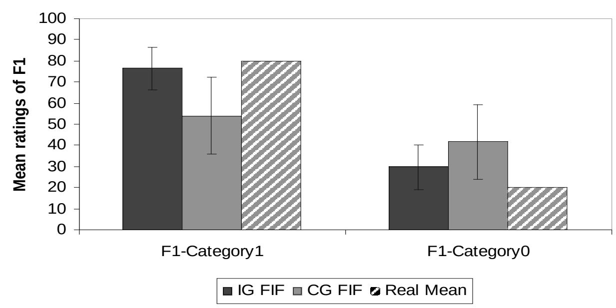

Figure 9: Goal effects in Fixed Inference format for inference test phase performance

Within the Classification format conditions, we also found a positive effect of consistency between the presentation format and the goal. For that format, one would expect consistency effects to arise in the classification test phase performance. The <Classification format, Classification goal> condition yielded more accurate classification ratings than the <Classification format, Inference goal> condition in 9 out of 10 cases, which is significantly better than by chance (p = .011, chi-square test).

In sharp contrast to these results, we found *no* format-goal consistency effect within the Random Inference format. <Random Inference format, Inference goal> and <Random Inference format, Classification goal> lead to similar inference test phase performance on all 4 features for both categories. We conjecture that task difficulty is playing a key role here. When the learning task is relatively easy (i.e., in Fixed Inference or Classification format), people are able to attend to the goal, and so inconsistencies between goal and presentation can matter. In contrast, when the mental load is heavier (i.e., in Random Inference format), people are sufficiently engaged in the learning task that they do not attend closely to the stated goal, and so do not notice the inconsistency. An alternative explanation is that all RIF participants had to memorize most of the exemplars to achieve 90% correctness threshold. If success is achieved through brute-force memorization of full cases, then different goals might not play a role. These explanations are not mutually exclusive, and substantially more evidence about precise cognitive loads and learning strategy is required to confirm these conjectures.

Conclusions

Because there are close connections between presentation format and task difficulty, it can be quite difficult to determine the source of differences in learned category representations. We offer here a novel type of presentation format—Fixed Inference format—that aims to remove between-format differences in task difficulty that are not due to intrinsic differences between inference and classification learning formats. We also aimed to determine whether participants' belief about subsequent category use played a role in category learning.

The experimental results suggest that previous findings of differential task difficulty might have been due to factors other than intrinsic differences between the learning formats. Factors such as variations in memory load may have played an important role, since the Fixed Inference format was not significantly more difficult than the Classification format on either measure of task difficulty. That being said, both analyses found the Fixed Inference format to be somewhat more challenging; we do not know whether the difference would prove to be statistically significant given a much larger sample size.

We also found that both presentation format and beliefs about subsequent category use are relevant for the learned category representation. The effects are (in some ways) not as dramatic as have been reported previously in the literature. The artificiality of our particular experiment stimuli might be at least partly responsible for differences in the reported magnitude of format effects.

Perhaps the most intriguing results are the findings of format-goal consistency effects for Classification and Fixed Inference formats, but not for the Random Inference format. Format-goal consistency has been studied in the educational psychology literature, but has received relatively little attention in concept learning. The present findings are particularly interesting since the consistency effects seemingly emerge only in "easier" conditions (though this is obviously speculative). We are currently investigating the nature and conditions under which these consistency effects robustly emerge.

Acknowledgments

Teddy Seidenfeld provided valuable conversations throughout the design and analysis process. D. Danks was partially supported by grants from NASA, ONR, and NIH. Three anonymous reviewers provided valuable comments.

References

Anderson, A. L., Ross, B. H., & Chin-Parker, S. (2002). A further investigation of category learning by inference. *Memory and Cognition, 30*, 119-128

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409-429.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory and Cognition, 30*, 353-362

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: a comparison of inference learning and classification learning. Journal of Experimental psychology: *Learning, Memory, and Cognition. 30*, 216-226

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

Murphy, G. L. (1993). A rational theory of concepts. In G. Nakamura, R. Taraban, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 29, pp. 327-359). Orlando, FL: Academic Press.

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Nilsson, H., & Olsson, H. (2005). Categorization vs. inference: Shift in attention or in representation? In B. G. Bara, L. Barsalou, & M. Bucciarelli, eds. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*. Mahwah, N.J.: Lawrence Erlbaum Associates. 1642-1647.

Ross, B. H. (1996). Category representations and the effects of interacting with instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1249-1265.

Ross, B. H. (1997). The use of categories affects classification. Journal of *Memory and Language, 37*, 1240-1267.

Ross, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 25*, 1743-1757.

Solomon, K. O., Medin, D. L., and Lynch, E. (1999). Concepts do more than categorize. *Trends in Cognitive Sciences, 3,* 99-105.

Yamauchi, T., & Markman, A.B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124-148.