# What Varying the Learning Task and Category Structure Reveals About Inference Learning

**Seth Chin-Parker (chinparkers@denison.edu)**
Department of Psychology, Denison University
Granville, OH 43023 USA

**Abstract**

A core issue in the cognitive sciences is understanding how people acquire conceptual knowledge. One way that people can acquire this knowledge is through the inference of missing feature information. Recent studies have proposed a shift away from the idea that inference learning results in knowledge of the internal structure of the categories being learned. The current study varies the inference learning task and the category structure being learned in order to examine these claims. The results provide little support for the notion that participants are acquiring either exemplar knowledge or a simple set of rules as a result of inference learning.

**Keywords:** categories and concepts, inference learning.

During the past decade, inference learning has been studied as a form of category learning (e.g. Sweller & Hayes, 2010; Yamauchi & Markman, 1998). Inference learning occurs as a participant predicts a missing feature value of an item when the category membership is explicitly available. For instance, the participant is shown a fictional bug, identified as a "DEEGER", that is missing its legs. Possible values for the missing legs are provided, and the participant predicts which would occur with that bug. As the participant continues to make inferences of this sort and receives feedback on those predictions, she acquires knowledge of the categories the bugs are drawn from.

Most previous research in this area has focused on comparing inference learning to classification learning. The current study focuses specifically on inference learning in order gain insight into that learning paradigm. One view of category learning is that the category knowledge acquired reflects that information about the category and its members that allows for a successful completion of the learning task (Markman & Ross, 2003). According to this view, inference learning leads to the acquisition of information about how features occur within the categories of interest. This can be as simple as learning the most likely feature value given the category label. However, this knowledge can also include more rich information about the internal structure of the category, e.g. the co-occurrence of feature values (Chin-Parker & Ross, 2002), variation among the feature values (Yamauchi & Markman, 2000), or abstract relations that exist between the features (Erickson, Chin-Parker, & Ross, 2005). Successful inference learning depends on realizing the category structure so that feature inferences can reflect how the features are instantiated within the category members. Although Johansen and Kruschke (2005) echo the notion that different learning tasks can lead to the acquisition of different category knowledge, they propose that inference learning results in a set of rules specifying only the association between the category label and the inferred feature values as opposed to a richer sense of the internal structure of the categories. More recently, Sweller and Hayes (2010) have proposed that inference learning results in the acquisition of exemplar knowledge under certain conditions, a claim that runs counter to much of the previous work examining inference learning. Although exemplar knowledge can be useful for many category-based tasks, inference learning seems to require a more unitary representation of the category to be successful.

The current study explores these claims by varying both the inference task and the structure of the categories being learned. Participants learned about categories that had a family resemblance (FR) structure. They either inferred only the prototypical feature values or both the prototype-consistent feature values and exceptions to those values. This manipulation has been used in previous studies (Nilsson & Olsson, 2005; Sweller & Hayes, 2010), and it has been proposed that inferring all feature values leads to exemplar learning. However, the effect of this manipulation on inference learning has not been fully explored. In this study, the participants also either learned a category structure where the feature values were shared across the categories or one where the feature values associated with each category were independent of one another. By examining the effect of the category structure, we can gain some insight into what knowledge is acquired through inference learning and how it is represented.

The working hypotheses for this study are based on the premise that successful inference learning involves selecting the most likely feature value in terms of the internal structure of the categories as opposed to a simple set of rules or exemplars. Several dependent measures are examined, including inference accuracy during learning and typicality ratings of both studied and novel items after the learning. Previous research (e.g. Sweller & Hayes, 2010) has shown that the manipulation of the inference task affects inference accuracy during learning. As noted prior, the inference task prompts the learner to select the most likely feature value given the category structure – this strategy leads to accurate inference of the prototype-consistent feature values but leads to poor accuracy when prototype-exception values are predicted. This in turn affects what participants are able to learn about the categories. Because their learning task better matches the category structure, the typicality ratings of the participants who infer only the prototype-consistent feature values will better reflect the FR structure of the categories. The effect of the category structure on participant performance will be more complex. Due to the nature of inference task, it is not obvious that the category structure manipulation will affect accuracy during learning because both category structures instantiate the prototypical feature values the same within the FR structure. However, I do predict that the participants who infer the feature values that are not shared across the categories will better realize the FR structure because they can focus on how specific feature values are distributed within each category. The participants who interact with the categories that share feature values will show less learning of the FR structure because they would have to realize the distribution of the feature values both within and across the categories to fully appreciate the category structure. What these results show about the underlying representation of category knowledge acquired from inference learning will be addressed in the Discussion.

## Experiment

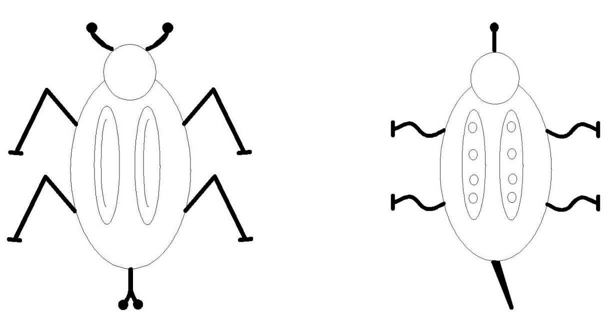
### Methods

**Participants** Ninety-six undergraduates from a Midwest college received participation credit for an Introductory Psychology class in return for their participation.

**Design** The experiment was a 2 (category structure) X 2 (inference task) independent samples design. During the learning task, the participants learned a category structure in which the feature values occurred across the categories, the *cross-category structure*, or the feature values were specific to each category, the *independent-category structure*. The structures of the categories learned are explained further in the Materials section of this paper. Also, participants either inferred both prototype-consistent and prototype-exception feature values (*exception inference conditions*) or only the prototype-consistent feature values (*consistent inference* conditions). The design resulted in four learning conditions: the cross-exception condition (C-E), the cross-consistent condition (C-C), the independent-exception condition (I-E), and the independent-consistent condition (I-C). Participants were randomly placed into a learning condition upon arrival at the experimental session.

**Materials** The stimuli were drawings of "bugs", labeled Deegers and Koozles (see Figure 1). The bugs varied along four features: the legs, antenna, tail, and wings. The stimuli

*Figure 1: Example Stimuli – Category Prototypes*



DEEGER (1111) KOOZLE (0000)

*Table 1: Category Structures Used in the Experiment*

Cross-Category Structure

| | Deeger | Koozle |
|---|---|---|
| Learning Items | 1110 | 0001 |
| | 1101 | 0010 |
| | 1011 | 0100 |
| | 0111 | 1000 |
| Prototype | 1111 | 0000 |

Independent-Category Structure

| | Deeger | Koozle |
|---|---|---|
| Learning Items | 1112 | 0003 |
| | 1121 | 0030 |
| | 1211 | 0300 |
| | 2111 | 3000 |
| Prototype | 1111 | 0000 |

were pretested to confirm that the feature values carried similar weights when participants were asked to make similarity judgments.

The categories (see Table 1) were defined by prototype, and during the learning task the participants interacted with bugs that matched the prototype on three of the four feature values (the *prototype-consistent* features), but had one feature value that did not match the prototype (the *prototype-exception* feature). The value of the prototype-exception feature depended on the category structure. In the cross-category structure, the values for prototype-exception features were the values associated with the prototype of the other category (e.g. a value of 0 for the prototype-exception feature in the Deeger category). In the independent-category structure, there was no overlap in the feature values associated with the two categories. For the learning stimuli, the feature that was to be inferred was removed from the picture of the bug, and pictures of each feature value were prepared in isolation so they could be displayed alongside the incomplete bug during the learning task as described in the Procedures section. Sixteen blocks of the learning exemplars were constructed so that during each learning block the participant would interact with each item once and each feature was inferred once for each category.

The items for typicality-rating task were based on the FR structure in Table 1. Participants rated the prototype of each category, *1-off items*, *2-off items*, *3-off items*, and *category conflict items*. The values of the prototype-exception features of these items varied as to whether they were from the cross-category, e.g. a value of 0 for a Deeger, or from the independent-category, e.g. a value of 2 for a Deeger. The 1-off items did not match the category prototype on one feature value, the 2-off items did not match on two feature values, and the 3-off items did not match on three feature values. The category conflict items had no prototype-consistent feature values (e.g. items 0000 and 2222 for the Deeger category). There were four 1-off items, six 2-off items, four 3-off items, and one category conflict item for each of the types of mismatch feature values, whether cross-category or independent-category values. Including the category prototype, there were 31 items from each category in the typicality-rating task. The stimuli for the single-feature classification task consisted of each of the feature values instantiated alone on the generic head-body form used for all bugs in the study.

**Procedure** The participants worked individually at a computer, seated roughly two feet from a 17" CRT monitor. Participants were told that they would be asked a series of questions about fictional bugs during the study. After this initial introduction to the study, instructions and reminders were presented on the computer.

The order of the sixteen learning blocks was randomized for each participant, as was the order of the presentation of the learning items within each block. For each trial of the learning task, a fixation point was followed by one of the learning items centered on the screen. As described in the Materials section, the bug was missing one of four features (the legs, wings, antenna, or tail). A notification of the category membership for the bug, either "This bug is a DEEGER" or "This bug is a KOOZLE", was presented in large font above the bug. To the right of the bug were the two possible values for the missing feature, the prototype-consistent value and the prototype-exception value. The location of each of the features, whether above or below the vertical center of the screen, was randomly determined each trial. Once the participant clicked on one of the feature values using the computer mouse, the initial picture of the bug was replaced with a complete version, all features were shown, and the participant was provided feedback on their inference. Either "CORRECT" appeared in a green font or "INCORRECT" appeared in a red font on each side of the bug for two seconds. The image of the bug remained for two seconds after the feedback before the next trial began.

After completing the 16 learning blocks, the participant was provided instructions for the typicality-rating task. The items seen during the task were blocked by category, and the order of the items within each block was random. The order of the categories was balanced across the participants. As in the learning, each trial was preceded by a short fixation point. The bug was presented in the center of the screen with the question, "How typical is this bug of a (DEEGER/KOOZLE)?" below it. Below the question was a seven-point typicality rating scale that was anchored at 1 ("Not at all typical"), 4 ("Somewhat typical"), and 7 ("Very typical"). The participant clicked on their rating using the computer mouse. Once the participant rated the 31 items in one of the category blocks, the other category block followed. No feedback was given during the task.

The final transfer task was the single-feature classification task. During this task, the feature values that were seen during the learning task were presented individually in a random order. Each feature value was preceded by a fixation point and was centered on the screen. On each side of the image were the category labels, "DEEGER" and "KOOZLE". The participant clicked on her classification for each feature value using the computer mouse. No feedback was provided during this task.

## Results

Figure 2 shows the mean learning performance for each of the conditions organized by the learning block quartiles. A 2 (category structure) X 2 (inference task) X 4 (learning block quartile) mixed ANOVA showed a significant main effect of learning block quartile, $F(3, 276) = 43.23$, $p < .001$, $\eta_p^2 = 0.32$, no main effect of category structure, $F(1, 92) = 0.00$, $p > .50$, and a main effect of the inference task, $F(1, 92) = 186.35$, $p < .01$, $\eta_p^2 = 0.67$. There was an interaction between the learning block quartile and inference task, $F(3, 276) = 13.58$, $p < .001$, $\eta_p^2 = 0.13$. As can be seen in Figure 2, the consistent-inference participants improved in their performance while the exception-inference participants showed much less improvement across the learning blocks. There was no effect of the category structure on this pattern. In the final learning block, 34 of the 48 consistent-inference participants were perfect at inferring the missing feature value, while only 4 of the 48 exception-inference participants did so. In contrast, 24 of the 48 exception-inference participants made four or fewer correct inferences in the final learning block.

Table 2 presents the mean typicality ratings for each of the item types by condition. It is important to note that all participants rated the typicality of items that had exception feature values seen during learning and items that had exception feature values not seen during learning. The analyses and tables that follow are organized to reflect this.

The analyses of the typicality ratings first focus on the *typicality slope* - the slope and intercept of the line that best fit each participant's ratings of the items from prototype to the conflict item. A near zero slope indicates that varying the number of prototype-consistent feature values has little effect on the typicality rating. A positive slope indicates that rated typicality increased as fewer feature values match the prototype (not expected if the FR structure is learned). A negative slope reflects that the participant is rating items

*Figure 2: Learning Performance by Quartile*

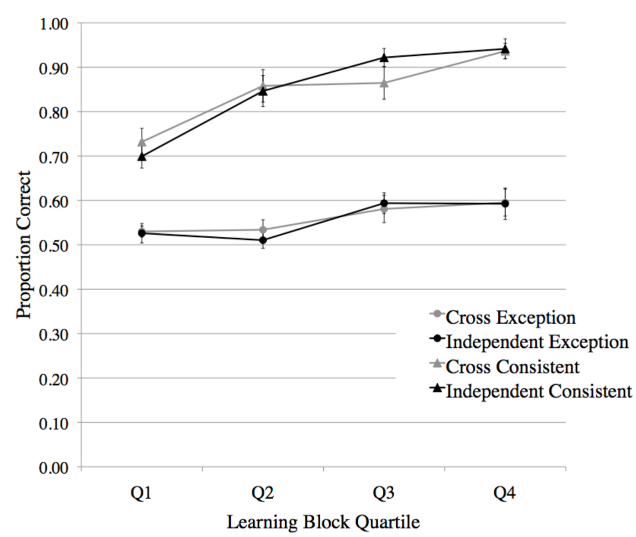



*Figure Note: Error bars represent +/- 1 SE*

*Table 2: Mean (and SD) of Typicality Ratings by Item*

*2a: Exception Feature Values Match Learning*

| | Item Type | | | | |
|---|---|---|---|---|---|
| | Proto | 1-0ff | 2-0ff | 3-Off | Conflict |
| C-E | 5.21 | 5.18 | 4.40 | 4.18 | 3.94 |
| | *(1.80)* | *(0.96)* | *(1.34)* | *(1.43)* | *(1.86)* |
| I-E | 5.15 | 5.45 | 4.80 | 4.51 | 4.29 |
| | *(1.68)* | *(0.96)* | *(1.26)* | *(1.40)* | *(1.82)* |
| C-C | 5.96 | 5.20 | 3.86 | 2.86 | 2.13 |
| | *(1.63)* | *(0.81)* | *(1.02)* | *(1.45)* | *(1.78)* |
| I-C | 6.08 | 5.47 | 4.98 | 4.49 | 4.29 |
| | *(1.43)* | *(0.92)* | *(1.13)* | *(1.53)* | *(1.79)* |

*2b: Exception Feature Values Mismatch Learning*

| | Item Type | | | | |
|---|---|---|---|---|---|
| | Proto | 1-0ff | 2-0ff | 3-Off | Conflict |
| C-E | 5.21 | 2.53 | 1.97 | 1.72 | 1.67 |
| | *(1.80)* | *(1.40)* | *(1.04)* | *(0.95)* | *(1.14)* |
| I-E | 5.15 | 3.80 | 3.14 | 2.72 | 2.77 |
| | *(1.68)* | *(1.18)* | *(0.95)* | *(0.92)* | *(1.74)* |
| C-C | 5.96 | 3.79 | 2.61 | 1.97 | 1.48 |
| | *(1.63)* | *(1.82)* | *(1.15)* | *(0.89)* | *(0.87)* |
| I-C | 6.08 | 4.49 | 3.45 | 2.77 | 2.25 |
| | *(1.43)* | *(1.07)* | *(0.98)* | *(1.22)* | *(1.45)* |

*Note: The prototype ratings are reported twice to allow for comparison across all levels in the tables above.*

that share fewer feature values with the prototype as less good members of the category. The slope intercept provides an indication of the typicality rating assigned to the more typical members of the category. The mean slope and slope intercepts were computed across the items that had prototype-exception values seen during learning (*match slopes*) and those that introduced the prototype-exception values not seen during learning (*mismatch slopes*). These values are reported in Table 3. The slope intercepts are not addressed here because of space.

The match slopes for all participants were analyzed using a 2 (category structure) X 2 (inference task) ANOVA. This analysis showed a main effect of category structure, $F(1, 92) = 8.93$, $p < .01$, $\eta_p^2 = 0.09$, a main effect of inference task, $F(1, 92) = 15.60$, $p < .01$, $\eta_p^2 = 0.15$, and an interaction between category structure and inference task, $F(1, 92) = 4.62$, $p = .03$, $\eta_p^2 = 0.05$. The participants that had been exposed to the cross-category structure during learning tended to have more negative match slopes than those exposed to the independent-category structure, and the participants in the consistent-inference conditions had more negative match slopes than those in the exception-inference conditions. The participants in the C-C condition had a mean match slope that was twice as large as any other condition leading to the interaction.

The mismatch slopes were similarly analyzed. A 2 (structure) X 2 (inference) ANOVA revealed a non-

*Table 3: Mean (and SD) Slope and Slope Intercept Values*

| | Exception Values Match Learning | | Exception Values Mismatch Learning | |
|---|---|---|---|---|
| | Slope | Slope Intercept | Slope | Slope Intercept |
| C-E | - 0.35 | 5.28 | - 0.79 | 4.20 |
| | *(0.53)* | *(1.31)* | *(0.52)* | *(1.20)* |
| I-E | - 0.26 | 5.36 | - 0.58 | 4.68 |
| | *(0.44)* | *(1.22)* | *(0.58)* | *(1.46)* |
| C-C | - 1.00 | 6.00 | - 1.07 | 5.31 |
| | *(0.62)* | *(1.13)* | *(0.42)* | *(1.70)* |
| I-C | - 0.46 | 5.97 | - 0.94 | 5.69 |
| | *(0.46)* | *(1.10)* | *(0.55)* | *(1.10)* |

significant effect of structure, $F(1, 92) = 2.63$, $p = .11$, $\eta_p^2 = 0.03$, a main effect of inference, $F(1, 92) = 9.28$, $p < .01$, $\eta_p^2 = 0.09$, and no interaction between the factors, $F(1, 92) = 0.10$, $p = .75$. The participants in the consistent-inference conditions had higher mismatch slopes on average than the participants in the exception-inference conditions.

The assessment of the typicality slopes is useful, but the slope summarizes the change across the levels of typicality. In order to look at the changes in the ratings at a more fine-grained level, I calculated for each participant how much the typicality ratings changed from the prototype to the 1-off items, the 1-off to 2-off items, etc. This provided four "drop" values for each participant. Each drop value indicated how the typicality ratings changed as an additional feature value mismatched the prototype. Using the four drop values, each participant's performance was categorized: *consistent drop* meant at least three of the four values were greater than 0.25, *consistent reversal* meant that three of the four values were less than -0.25, *flat* meant that three of the four values were between -0.25 and 0.25, and *inconsistent* meant that the drops varied across these ranges. Table 4 shows these patterns organized by condition and whether the values of the exception features matched (Table 4a) or mismatched (Table 4b) those seen during the learning. Organizing the patterns into two types, *Consistent* and *Other*[1], there is a obvious effect of the condition, $\chi^2_{(3,\ 96)} = 13.72$, $p < .01$, $\phi_c = 0.38$. Only the C-C condition had a majority of the participants showing a consistent drop across these items in the typicality rating task when the prototype-exception feature values matched those seen during learning. There is also an effect of condition on the patterns observed in the typicality ratings for the items with the mismatched exception feature values, $\chi^2_{(3,\ 96)} = 11.31$, $p = .01$, $\phi_c = 0.34$. The typicality ratings of the C-E condition did not consistently reflect the typicality gradient. The I-E condition was more consistent in their ratings. Most of the consistent-inference participants showed a consistent change in their typicality ratings. It is important to note that nearly

[1] Combining the reversal, flat, and inconsistent into one grouping to contrast with consistent is necessary for the analysis because of the small number of participants with the flat and reversal patterns.

*Table 4: Distribution of Participant Typicality Drop Patterns Within the Conditions*

*4a: Exception Feature Values Match Learning*

| | Consistent | Reversal | Flat | Inconsistent |
|---|---|---|---|---|
| C-E | 9 | 1 | 2 | 12 |
| I-E | 6 | 2 | 4 | 12 |
| C-C | 18 | 1 | 1 | 4 |
| I-C | 9 | 0 | 3 | 12 |

*4b: Exception Feature Values Mismatch Learning*

| | Consistent | Reversal | Flat | Inconsistent |
|---|---|---|---|---|
| C-E | 8 | 1 | 8 | 7 |
| I-E | 13 | 1 | 3 | 7 |
| C-C | 16 | 0 | 4 | 4 |
| I-C | 19 | 0 | 0 | 5 |

all participants who were categorized as having a flat response pattern with the mismatch items (especially those in the C-E condition) actually rated the prototype rather high but the introduction of any feature value not seen during learning resulted in a low typicality rating.

The final measure was the single feature classification accuracy. It is important to remember that in this task the cross-category conditions (C-E and I-E) classified eight feature values while the independent-category conditions (I-E and I-C) classified sixteen feature values. All four conditions, C-E (M = 0.79, SD = 0.14), I-E ($M$ = 0.87, $SD$ = 0.15), C-C ($M$ = 0.89, $SD$ = 0.19), and I-C ($M$ = 0.86, $SD$ = 0.14), were above chance when classifying the feature values (all $p$ < .01). A 2 (category structure) X 2 (inference task) ANOVA showed no effect of category structure, $F(1, 92) = 0.89$, $p$ = .35, $\eta_p^2 = 0.01$, no effect of inference task, $F(1, 92) = 1.84$, $p$ = .18, $\eta_p^2 = 0.02$, and a marginally significant interaction between category structure and inference task, $F(1, 92) = 3.12$, $p$ = .08, $\eta_p^2 = 0.03$. The interaction term approaches significance, reflecting the similar performance of all conditions except for the C-E condition. Both the C-C and I-E conditions were significantly more accurate than the C-E condition (both $p$s < .05), and the difference between the C-E and I-C conditions approached significance ($p$ = .07).

## Discussion

The results of this experiment provide some insight into the nature of inference learning. To facilitate discussion, a brief summary of the experimental conditions follows. The cross-exception (C-E) condition performed poorly during the learning task, and their typicality ratings showed restricted knowledge of the FR category structure. When the exception values matched those seen during learning, the typicality slope was small and inconsistent across the participants in the condition. When rating the mismatch items, the participants in the C-E condition gave any item with a novel feature value a low rating; the number of prototypical values had little impact on their ratings. The independent-exception (I-E) condition, like the C-E, had difficulty during the learning. Their typicality ratings for the matched items were also similar to the other exception-inference condition. However, when they were rating the typicality of the mismatch items, they showed more sensitivity to the FR structure. The cross-consistent (C-C) condition performed very well during the learning. Their typicality ratings showed that they were sensitive to the FR structure of the categories they learned, and this was evident whether they were rating items that had exception features that matched those that they had seen during learning or not. The independent-consistent (I-C) condition also did well during the learning task. In the matched slope analyses, the typicality slope of the I-C condition was somewhat flat and the participants' ratings fluctuated among the items. However, the ratings of the mismatch items, in terms of both the slope measures and the consistency of the participant drops showed recognition of the FR structure.

As predicted, there was a strong effect of the inference task, but little effect of the category structure, during the learning. The lower learning performance of the exception-inference conditions occurred because the participants consistently predicted the prototypical value for the missing feature. On 25% of the learning trials, the participants were getting feedback that the prototypical value was the incorrect value for the missing feature. As noted prior, this works directly against the underlying principle of the inference task – identifying the most likely, i.e. prototypical, feature value given the category of interest. This disruption in the learning was extensive; half of the participants in the exception-inference condition correctly predicted four or fewer of the missing features in the final block of learning.

The effect of the inference task on the transfer measures is more complicated, and it interacts with the effect of the category structure. Within the cross-category conditions, the learning task had a strong effect. Participants in the C-E condition did not realize the category structure, but the C-C condition showed evidence across all of the transfer tasks that they understood the FR structure. The independent-structure conditions were more similar, although more variable, in terms of what the typicality ratings suggested about their knowledge of the categories. The typicality ratings for the items with exception-feature values that matched those seen during learning indicated that participants in both the I-E and I-C conditions did not appreciate the FR structure. The ratings for the items with the mismatch values indicated that they did understand the FR structure, the I-C condition somewhat better than the I-E condition. The difference between the ratings for the items with the matched and mismatched exception-feature values can be attributed to two factors. First, within this category structure all feature values in the matched items had been associated with a single category during learning, and this may have attenuated the drop in the ratings for the less typical items. Second, the items with the matched and mismatched exception-feature values were interleaved, so the ratings may have been treated as relative – regardless of the number of prototypical features, the matched items were still more typical than the mismatched items. Importantly, both independent-category conditions showed evidence of recognizing the FR structure to some degree following the learning. The predicted overall advantage for the independent category structure was not seen as the C-C condition learned the categories as well as either of the independent-structure conditions. The preservation of some learning in the I-E condition, especially compared to the C-E condition, is the sole indication that the independent category structure affected learning as predicted.

As noted, these results can be used to address recent proposals about inference learning. First, Johansen and Kruschke (2005) proposed that inference learning leads to a set of rules representation that specifies the associations between the category label and feature values inferred during learning. The simplest form of the model can be ruled out as the I-C condition showed sensitivity to the value of the exception features during the typicality rating task. This indicates that some information about the non-inferred feature values was captured in the category representation. Johansen and Kruschke acknowledged that a more complex rule model might be necessary to capture the knowledge acquired from the inference learning. Indeed, the set of rules would have to somehow capture more than just the inferred and non-inferred feature values; it would also have to be flexible enough to represent relationships between feature values (Chin-Parker & Ross, 2002) and abstract relations between features (Erickson, Chin-Parker, & Ross, 2005).

This study found very little to support the claim that the exception-inference task results successful storage of exemplar information during learning. As in prior studies using the exception-inference task and cross-category structure (Nilsson & Olsson, 2005; Sweller & Hayes, 2010), participants performed poorly during the learning task. Here, details about the progression of the inference accuracy during the learning trials are provided so that we can better appreciate the difficulty of the exception-inference task. If participants had based their predictions on individual exemplars instead of predicting the most likely value given the category, performance would improve. However, only five of the forty-eight participants in the current exception-inference conditions were above 75% accuracy in the final learning block. So, although it is possible that participants could use exemplar knowledge to guide their inferences, it was not a strategy that was readily employed in this experiment. Also, a review of the transfer measures from this study and the others shows that exception-inference participants perform poorly on tasks designed to illustrate their category knowledge, especially when feature values are distributed across categories, complicating attempts to ascertain the representation that underlies that knowledge.

The evidence suggests that the exception-inference task undermines a fundamental constraint of inference learning – a stable relationship between the category label and feature values so that the learner can develop a coherent sense of the internal structure of the category. Sweller & Hayes (2010) are correct that exceptions to prototypical features exist, but they may be learned through a process other than direct inference. Importantly, the current study provides evidence that these constraints may be moderated by other factors like the category structure being learned.

The pattern of results reported here fits with the notion that inference learning relies on the development of knowledge of the internal structure of the categories. When the focus of the inference task matched the FR structure, as with the consistent-inference conditions, participants easily learned the categories. When there was a mismatch, e.g. the C-E condition, learning was meaningfully hindered. Current research, e.g. Yamauchi (2009), is exploring possible processes that underlie the abstraction process that occurs during feature inference and how coherent representations of the category structure develop.

## Acknowledgments

Avi Baranes was invaluable in his help with this study. He assisted in development of the stimuli and data collection.

## References

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & 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, & Cognition, 30,* 216–226.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 1433-1458.

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

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 Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Sweller, N., & Hayes, B. K. (2010). More than one kind of inference: Re-examining what's learned in feature inference and classification. *The Quarterly Journal of Experimental Psychology, 63,* 1568-1589.

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

Yamauchi, T., & Markman, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition, 28,* 64–78.

Yamauchi, T. (2009). Finding abstract commonalities of category members. *Journal of Experimental & Theoretical Artificial Intelligence, 21,* 155-180.