Costly Exceptions: Deviant Exemplars Reduce Category Compression

Daniel C. Silliman (dsillim1@binghamton.edu
Sean Snoddy (ssnoddy1@binghamton.edu)
Matthew Wetzel (mwetzel2@binghamton.edu)
Kenneth J. Kurtz (kkurtz@binghamton.edu)

Department of Psychology, Binghamton University (SUNY)
Binghamton, NY 13902 USA

Abstract

We investigated whether the presence of exception items can impede effects of category compression (within-category items appearing more similar) in classification learning. We hypothesized that the distinct representations afforded to exceptions may cause the target category to appear less cohesive, thereby reducing the likelihood of compression occurring. Across two experiments, participants engaged in classification learning without exceptions, with an easy exception, or with a difficult exception. Pairwise similarity ratings for all items were collected before and after learning to index compression. Results from Experiment 1 suggest that difficult exceptions can impede compression for the contrast category when situated within its cluster, while results from Experiment 2 suggest that both kinds of exceptions can impair compression of standard items in a target category relative to the No Exception control. We also observed surprising evidence of a novel between-category compression effect that was observed with the category structure developed for these experiments.

Keywords: category learning; learned categorical perception; similarity; representation change; exceptions

Introduction

Category learning is widely considered to involve not only forming associations between items and classes, but simultaneously learning to represent the items in a manner that reflects their categorization (e.g., Goldstone, Lippa, & Shiffrin, 2001). Researchers interested in representation change as a consequence of learning have pursued two distinct lines of research: 1) studying changes in the representation of all training items; 2) studying changes in the representation of exception items that violate norms in the distributional structure of the categories.

The former line of research has revealed two possible changes to exemplar representation that can occur, collectively referred to as learned categorical perception (CP) effects. Exemplars within a common category are judged to be more similar to each other after learning—an effect referred to as *within-category compression* (simply referred to as compression hereafter)—and exemplars across different categories are judged to be less similar to each other—referred to as *between-category expansion*, or simply expansion (Goldstone et al., 2001; Kurtz, 1996; Livingston, Andrews, & Harnad, 1998).

These changes to perceived similarity are believed to be driven by stable changes to underlying representations (or re-weighting of features), rather than strategic judgements deviating from the actual representations or temporary task-specific commitments (Folstein, Palmeri, & Gauthier, 2013; Goldstone et al., 2001).

Relatedly, research examining changes to learned exception items also indexes representational differences. Unlike research on learned CP effects, however, research on exceptions focuses more on post-learning differences between the exception(s) and the standard items rather than broad-based shifts as a function of learning and task demands. This research relies on recognition tests to demonstrate that exceptions form distinct representations from standard items (e.g., Sakamoto & Love, 2004).

Despite these two literatures having overlap in aims and goals, there has been little to no research examining the possible interactions that might arise from learning representations for exceptions and standard items simultaneously. Tangentially related work has examined the influence of various category structures (Pothos & Reppa, 2014) on CP, but little work has closely examined the consequences of forming a distinct exception representation for CP effects.

The purpose of the present work is to investigate these consequences. Specifically, we aim to determine what influence exceptions have on compression and expansion effects in standard items. To these ends, two experiments were conducted wherein changes to exemplar pairwise similarity ratings were observed in relation to learning with two different types of exceptions (as well as a control of No Exception). In the Easy condition, the exception had minimal similarity to either category. In the Difficult condition, the exception had greater similarity to the contrast category than its own host category. The decision to use multiple types of exceptions was motivated by recent work suggesting that the aforementioned special status of exceptions may have less to do with their violating knowledge structures in general, and more to do with their similarity to the opposing category (Savic & Sloutsky, 2017). Assuming unique exception representations do affect compression and expansion, this relationship should be stronger in conditions that encourage unique representations (i.e., the Difficult condition).

Regarding how the exception items may influence CP—it is possible that in forming distinct exception representations,

the exception item may cause its host category to appear less cohesive overall. Despite having relatively low within-category similarity, the uniquely represented (and consequently disproportionately weighted) exception remains strongly associated with the category. A less cohesive category may be more difficult to compress. Depending on whether the exception is highly confusable with the contrast category, the contrast category may also appear less cohesive and, again, less susceptible to compression effects.

Experiment 1

We conducted a traditional classification learning task bookended by pairwise similarity ratings for all exemplars. The primary aim of this study was to determine if there are relative differences in CP effects across conditions. We also sought to replicate classic CP findings by checking for the presence of these baseline effects in each condition and category individually (CP effects were operationalized as significant difference from 0 in either direction).

For classification, we predicted lower accuracy for the difficult exception given its high confusability with the contrast category. For these same reasons, we predicted the difficult exception would be harder to integrate with its host category than the easy exception, resulting in reduced category cohesiveness. Consequently, the standard items in the Difficult condition should evidence less compression than those same items in the Easy and No Exception conditions. We also expected the difficult exception to cause the contrast category to appear less cohesive given that the exception resides in its cluster. As a result, the contrast category for the Difficult condition should also evidence less compression relative to the Easy and No Exception conditions.

Regarding individual CP effects, we expected to find traditional within-category compression and between-category expansion effects for the control (No Exception) condition. All other analyses of this type were exploratory.

Method

Participants Binghamton University undergraduates (N = 168) participated in this experiment. Three participants were dropped from analyses (leaving 165) for failing to follow task instructions.

Materials and Design The stimuli were comprised of squares that varied on dimensions of size and shading (see Figure 1). The stimuli values were selected as a subset from a larger set that has been previously demonstrated to provide equally salient dimensions. For the Easy and Difficult conditions, an equal number of squares were assigned to two separate categories defined by a diagonal structure—i.e., participants had to attend to both dimensions to learn the category properly (see Figure 1). For the exception conditions, only Category A included an exception.

In the control condition, there was one item fewer in Category A than Category B (due to the missing exception); this decision was made to hold the number and appearance of the standard items constant across the conditions. Both the difficult and easy exceptions were equidistant from the nearest member of the A Category, while the easy exception was also equidistant from the nearest member of both the A and B categories. We used a between-subjects design with assignment to condition randomized across participants.

Procedure Participants began with the first similarity rating phase. All possible pairwise combinations for a given condition were seen, with no pair repeating. The order of the pairs was randomized by participant. Each participant was instructed to use an on-screen, unmarked, continuous slider to indicate how visually similar the two on-screen squares were. Participants were informed of the two relevant dimensions and told to factor both into consideration when making their decision. A mouse click was used to indicate selected response. Although participants could not see the numerical values of the scale, they ranged from -50 to +50.

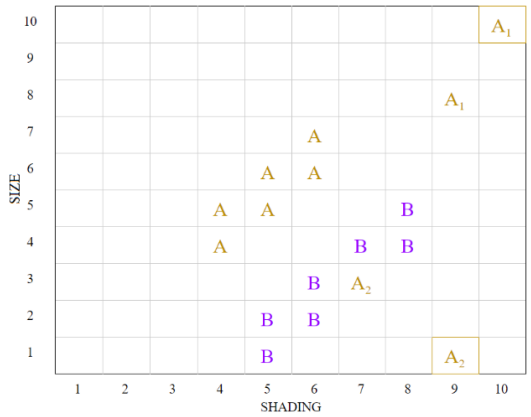

Figure 1: Illustration of category structure used in the Experiments. Subscripts indicate the additional items included in only the Easy (1) or Difficult (2) conditions. Outlined cells denote exception items for Experiment 2.

In the classification phase items were presented one at a time, centered on-screen with two category labels to choose from. Responses were made with mouse clicks. At the beginning of the phase, participants were informed that the squares belong to two categories, and that their task was to determine which belong to a given category. Feedback was provided for both incorrect and correct choices. Each block consisted of every exemplar presented in a randomized order. Each participant completed three blocks of learning. After the classification phase, the second rating phase began with similar instructions and identical procedures as the first.

Results & Discussion

Classification Training An initial question is whether exception items impacted classification accuracy for the standard items. A linear mixed effects regression (LMER) predicting accuracy from condition, block, and their interaction—and allowing participant to vary as random intercepts—was used to address this question. The Difficult

condition did not significantly differ from the Easy (β = -0.04, SE = 0.037, t = -0.1081, p = .28) or No Exception (β = -0.016, SE = 0.036, t = -0.451, p = .652) condition; the Easy condition did not significantly differ from the No Exception condition (β = 0.023, SE = 0.036, t = 0.646, p = .518) (Figure 2, left panel). Training block was a significant predictor of accuracy in the Difficult (β = 0.053, SE = 0.011, t = 4.971, p < .001), Easy (β =0.053, SE = 0.01, t = 5.123, p < .001), and No Exception (β = 0.062, SE = 0.01, t = 6.067, p < .001) conditions confirming that accuracy increased across training blocks. There were no significant interactions between condition and block (all ps > .529).

Learning of the exception items alone was analyzed via a logistic mixed effects regression using the same predictors as the previous model. There was no significant difference between conditions (β = 0.746, SE = 0.812, p = .358) and no effect of block (β = 0.161, SE = 0.284, p = .572). However, there was a significant interaction between condition and block (β = 1.19, SE = 0.402, p = .003), such that accuracy for the easy exception increased more during training than accuracy for the difficult exception (Figure 2, right panel). An exploratory binomial test revealed that the final block of classification performance in the difficult exception was significantly below chance performance (17%, N = 52, p < .001), suggesting that participants were unable to learn that the difficult exception belonged to Category A.

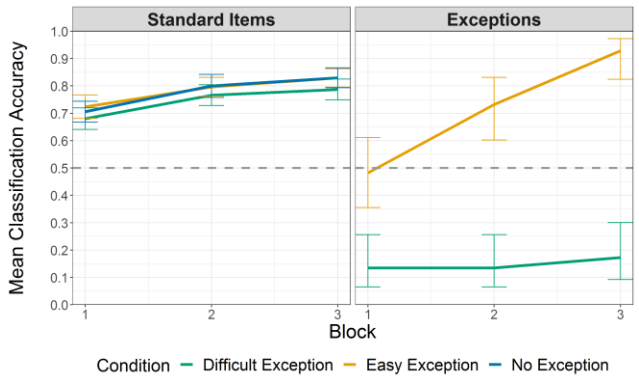

Figure 2: Mean classification for Experiment 1. Error bars represent 95% confidence intervals (Exceptions panel uses binomial confidence intervals). The dashed line reflects chance classification performance.

Although we predicted lower accuracy for the difficult exception, we did not anticipate that participants would completely fail to learn the item. This might be due to a combination of overly high confusability with the contrast category and a low frequency of exposure. Given that the majority of participants may have formed an indeterminate category association for this item at the end of learning, it is unlikely that the item formed a distinct representation as an A item (as is typically found in studies on exceptions). We address this in Experiment 2 by adjusting the nature and frequency of the exception.

Within-category Similarity Ratings To determine whether our design and materials replicate traditional baseline compression effects, a series of one sample t-tests were performed on the (post – pre) similarity rating score from the same-category pairs of items. Separate analyses were conducted for pairings including just the standard items and pairings including the exceptions. We separated analyses by item type to better gauge if any differences in CP effects reflected the influence of the exception on the whole category, rather than just the pairwise comparisons including the exceptions themselves (which may disproportionately affect the rating differences). These analyses were further separated by condition and, where appropriate, category (see Figure 3). Due to the number of t-tests conducted, a Bonferroni adjustment was made to the alpha level, resulting in a new alpha of .00625. Under these analyses, finding a significant positive mean difference from zero provides evidence of compression. With the exception of the standard Category B items of the Difficult condition (p = .073), and pairings including the difficult exception (p = .319), all other tests revealed significant positive differences from zero (all ps < .00625, Bonferroni adjusted). These results replicate the traditional compression effects and reveal potential new effects owed to the presence of exceptions.

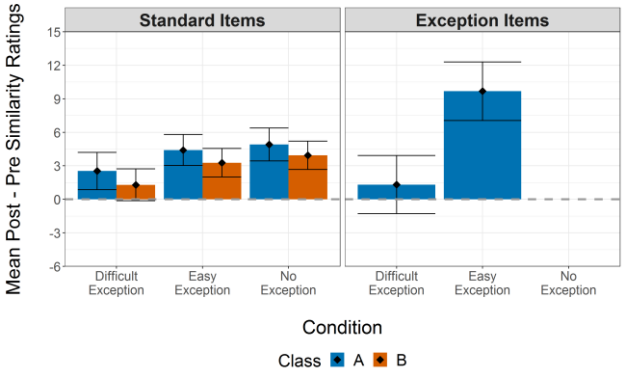

Figure 3: Within-category mean difference ratings for Experiment 1. Error bars reflect 95% CIs, diamonds reflect adjusted means obtained from the respective model. The dashed line represents no change in similarity ratings.

Our primary questions and predictions concerned relative CP differences across conditions. We predicted that the Difficult condition would evidence comparatively less compression than the Easy and No Exception conditions, due to the host category appearing comparatively less cohesive. To determine if this was the case, an LMER allowing participant to vary as random intercepts and predicting (post – pre) rating differences from condition, category, and their interaction was used to address these questions. For the standard items, the No Exception condition did not significantly differ from the Difficult (Category A as reference: β = 2.365, SE = 1.692, t = 1.397, p = .163; Category B as reference: β = 2.643, SE = 1.594, t = 1.658, p = .099) and Easy (Category A as reference: β = 0.492, SE = 1.661, t = 0.296, p = .767; Category B as reference: β = 0.668, SE =

1.564, t = 0.427, p = .67) conditions. The Easy condition did not significantly differ from the Difficult condition (Category A as reference: β = 1.872, SE = 1.7, t = 1.102, p = .272; Category B as reference: β = 1.875, SE = 1.601, t = 1.234, p = .219). With respect to the exception items, the Easy condition had significantly larger difference ratings than the Difficult condition (β = 8.373, SE = 2.666, t = 3.141, p = .002) (Figure 3, right panel).

Our predictions regarding mitigated compression in Category A of the Difficult condition relative to other conditions were not supported. Our ability to meaningfully interpret these data in light of our hypotheses is somewhat hampered by the unexpectedly poor learning of the difficult exception. We address this problem in Experiment 2. Despite this outcome, there is still one finding that is pertinent to our hypotheses. When investigating for the presence of baseline CP effects, we found that standard items in Category B of the Difficult condition did not evidence compression (defined here as a significant positive difference from zero). Though the baseline CP effects were intended as manipulation checks, they also speak to the primary aims of the study. The cohesiveness of category B under the Difficult condition was likely compromised by the presence of the difficult exception in its cluster. Despite appearing to be a B-item, participants would receive corrective feedback to the contrary (as supported by the low accuracy). Learning that not all items that appear as *B*s are actually *B*s likely affected participants' representation for that category such that it was less amenable to compression. Though there was no evidence of compression for this category under the Difficult condition, we also note that there are no significant differences between this category across the other conditions (which did evidence compression). This suggests that the influence of the difficult exception on compression may not be that considerable.

Regarding CP effects for the exception items themselves, we found that compression between the easy exception and Category A appears relatively high, though this is likely due to it appearing very different from all other items prior to learning. We note that it does not greatly affect compression in the standard items.

Between-category Similarity Ratings To determine whether baseline between-category expansion effects occurred in any of the conditions, a series of one sample t-tests were performed on the rating differences from between-category pairs. A Bonferroni adjustment was made, resulting in an alpha of .01. Surprisingly, none of the conditions demonstrated traditional between-category expansion effects. The standard items in Easy and No Exception conditions and the easy exception item all demonstrated mean ratings significantly above zero (all ps < .01), suggesting an unusual outcome of between-category compression. The Difficult standard items (p = .904) and exception (p = .053) demonstrated neither expansion nor compression effects (See Figure 4).

We next investigated relative expansion differences across conditions. No *a priori* predictions were made regarding potential differences. To determine whether between category rating differences were affected by condition, an LMER that predicted rating differences with condition and allowed participant to vary as random intercepts was used. For the standard items, there were no significant differences between the Difficult and No Exception condition (β = -2.096, SE = 1.742, t = -1.204, p = .23), Difficult and Easy conditions (β = -2.968, SE = 1.749, t = -1.697, p = .092), and Easy and No Exception conditions (β = 0.871, SE = 1.709, t = 0.51, p = .611) (Figure 4, left panel). For the exception items, the Easy condition lead to greater between category compression than the Difficult condition (β = 8.047, SE = 2.122, t = 3.792, p < .001) (Figure 4, right panel).

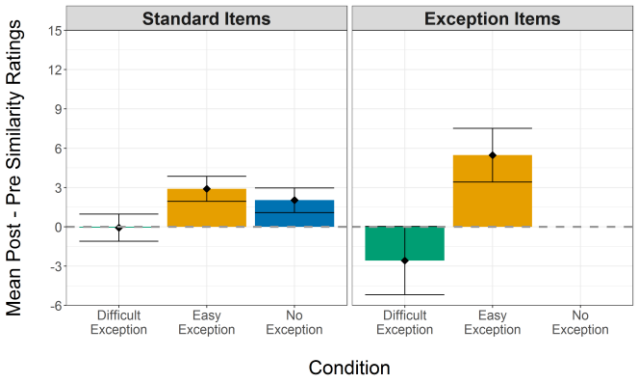

Figure 4: Between-category mean difference ratings for Experiment 1. Error bars reflect 95% CIs, diamonds reflect adjusted means obtained from the respective model. The dashed line represents no change in similarity ratings.

A surprising result of between-category compression was observed in both the Easy and No Exception conditions. We could not find this effect in our review of the CP literature involving category and concept learning. This finding may be due to the use of the diagonal structure with non-integral features, which requires attention to more than one feature for successful performance. It is possible that expansion is an artifact of category structures where a single dimension is sufficient to differentiate the categories. It should be noted that the majority of research for both CP and exception learning use structures where one dimensional solutions are sufficient for adequate performance (though see Pothos & Reppa, 2014). The integration of both dimensions in a representation may cause both categories to seem more similar overall. This effect may be absent in the Difficult condition due to the similarity of the exception to Category B. Curiously, however, learning in that condition was not significantly less than the other conditions, suggesting that it is not an impediment of learning that prevents the effect.

Experiment 2

Due to the unforeseen difficulty of learning the difficult exception in Experiment 1, the aim of Experiment 2 was to determine if greater exposure to both standard and exception items could result in more robust CP differences across conditions. In addition to increased training for all items, the

difficult exception was altered to make it easier to learn. A concomitant change to the easy exception was also made to roughly equate initial distances between both exceptions and their host category. We predicted that if participants learn the difficult exception, then Category A should appear less cohesive in light of the distinct exception representation, resulting in no evidence of compression for the standard items. An ancillary aim was to determine if the between-category compression observed in Experiment 1 would replicate with the modest changes to materials and design.

Method

Participants Binghamton University undergraduates (N = 156) participated in this experiment.

Materials and Design The items and the design were largely the same as in Experiment 1. The difficult exception was changed to coordinate [9,1] of Figure 1, while the easy exception was changed to coordinate [10,10].

Procedure Only two changes were made from the procedure of Experiment 1. First, the number of classification learning blocks was increased from three to six. Second, instead of presenting the exception item as frequently as standard items, the number of exposures per block was increased to three.

Results & Discussion

Classification Training To analyze learning performance, the same LMER model from Experiment 1 was used. In contrast to Experiment 1, the Difficult condition resulted in significantly lower classification accuracy than the Easy (β = -0.091, SE = 0.03, t = -3.07, p = .002) and No Exception (β = -0.131, SE = 0.029, t = -4.497, p < .001) conditions. The Easy condition did not significantly differ from the No Exception condition (β = 0.04, SE = 0.029, t = 1.407, p = .16) (Figure 5, left panel). Training block was a significant predictor of accuracy in the Difficult (β = 0.041, SE = 0.006, t = 7.418, p < .001), Easy (β =0.045, SE = 0.005, t = 8.198, p < .001), and No Exception (β = 0.049, SE = 0.005, t = 9.178, p < .001) conditions, such that accuracy increased across training blocks. There were no significant interactions between condition and block (all ps > .366).

To test whether the exception item differed between Difficult and Easy conditions, an LMER was used that predicted average block accuracy with condition, block, and their interaction and allowed participant to vary as a random intercept. The exception in the Easy condition was associated with significantly higher accuracy than in the Difficult condition (β = 0.411, SE = 0.051, t = 8.017, p < .001). There was a significant, positive effect of training block on accuracy (β = 0.084, SE = 0.01, t = 8.574, p < .001). There was a significant interaction between condition and block (β = -0.041, SE = 0.014, t = -3.02, p = .003), such that accuracy for the easy exception did not increase throughout training, while accuracy improved throughout training for the difficult exception (Figure 5, right panel).

The modifications for this experiment appear to have achieved the intended effect, given that the difficult exception was adequately learned, while overall performance was still less than that of the easy exception. Interestingly, the standard items for the Difficult condition also appear to have been learned less well than the Easy and No Exception conditions. Discovering the diagonal structure may be complicated by the repeated appearance of the difficult exception.

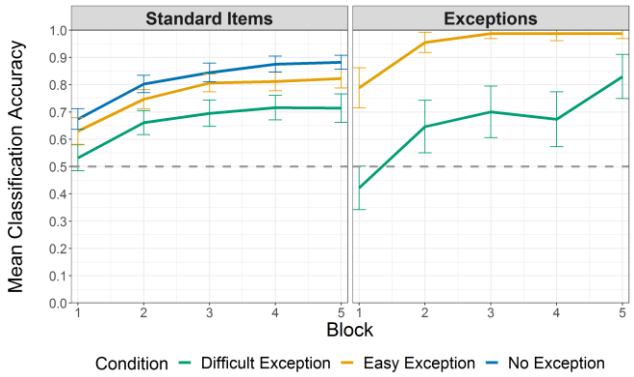

Figure 5: Mean classification accuracy for Experiment 2. Error bars represent 95% confidence intervals. The dashed line reflects chance classification performance

Within-category Similarity Ratings As with Experiment 1, a series of one sample t-tests were performed on the (post – pre) rating differences for within-category pairs of items (see Figure 6) to determine if traditional baseline compression effects were observed. The Bonferroni adjusted alpha was again set to .00625. The only items that did not significantly differ from zero were the standard items in Category A within the Difficult condition (p = .009). All other items were significantly higher than zero (all ps < .00625), which suggests that compression occurred.

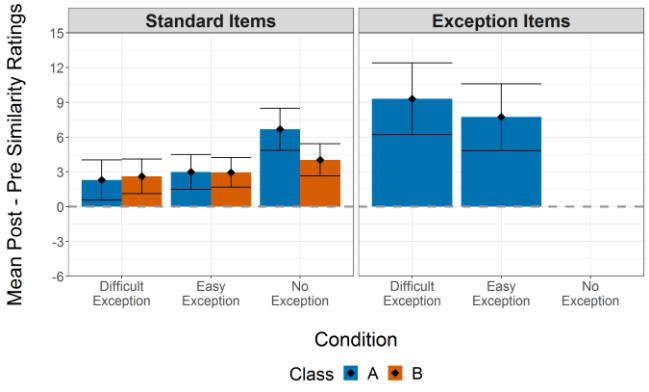

Figure 6: Within-category mean difference ratings for Experiment 2. Error bars reflect 95% CIs, diamonds reflect adjusted means obtained from the respective model. The dashed line represents no change in similarity ratings.

The primary question for within-category ratings was whether condition affected the observed compression effects.

An LMER that allowed participant to vary as random intercepts and predicted (post – pre) rating differences with condition, category, and their interaction was used to address this question. For the standard items, the No Exception resulted in significantly higher difference ratings than the Difficult (β = 4.365, *SE* = 1.78, *t* = 2.439, *p* = .0154) and Easy (β =3.675, *SE* = 1.762, *t* = 2.086, *p* = .038) conditions for Category A items, but did not differ from the Difficult (β = 1.428, *SE* = 1.681, *t* = 0.85, *p* = .397) and Easy (β = 1.088, *SE* = 1.655, *t* = 0.657, *p* = .512) conditions for Category B items. The Easy condition did not significantly differ from the Difficult condition the (Category A as reference: β = 0.68, *SE* = 1.81, *t* = 0.38, *p* = .704; Category B as reference: β = 0.34, *SE* = 1.704, *t* = 0.2, *p* = .842). With respect to the exception items, the Easy condition did not differ from the Difficult condition (β = 1.566, *SE* = 3.574, *t* = 0.438, *p* = .662) (Figure 6, right panel).

Our primary prediction was that the difficult exception would cause Category A to appear less cohesive, thereby reducing the degree of compression observed for the standard items in that condition. This prediction is borne out by the data. That this pattern absent in Experiment 1 is likely due to participants successfully learning the difficult exception in this experiment.

When comparing across conditions, both the Difficult and Easy condition evidenced less compression in Category A (associated with the exceptions) compared to the No Exception condition. This finding suggests that the presence of any exception is sufficient to impede compression effects to some degree. Whatever representation an exception does attain may be sufficient to compromise compression.

In contrast with Experiment 1, Category B of the Difficult condition is now evidencing compression. This may be due to altering the difficult exception so that it is less confusable with Category B, making it less likely that the exception impinges on the cohesiveness of the category.

Between-category Similarity Ratings The same series of one sample *t*-tests in Experiment 1 were performed on the rating differences from between category pairs. The alpha was again set to .01. As in Experiment 1, the standard items in Easy and No Exception conditions, including the easy exception item, all demonstrated between-category compression effects (all *p*s < .01, Bonferroni adjusted), while the standard items (*p* = .031) and exception (*p* = .733) in the Difficult condition demonstrated neither expansion nor compression effects (see Figure 7).

To determine if category rating differences were affected by condition, the same LMER structure used in Experiment 1 was employed. For the standard items, there were non-significant trends in favor of lower differences in the Difficult condition relative to the No Exception condition (β = -3.371, *SE* = 1.846, *t* = -1.826, *p* = .07) and Easy (β = -3.544, *SE* = 1.871, *t* = -1.894, *p* = .06) conditions. There was no significant difference in rating differences between the Easy and No Exception conditions (β = 0.173, *SE* = 1.818, *t* = 0.095, *p* = .924) (Figure 7, left panel). For the exception items, the Easy condition led to greater between category compression than the Difficult condition (β = 5.894, *SE* = 2.346, *t* = 2.513, *p* = .014) (Figure 7, right panel).

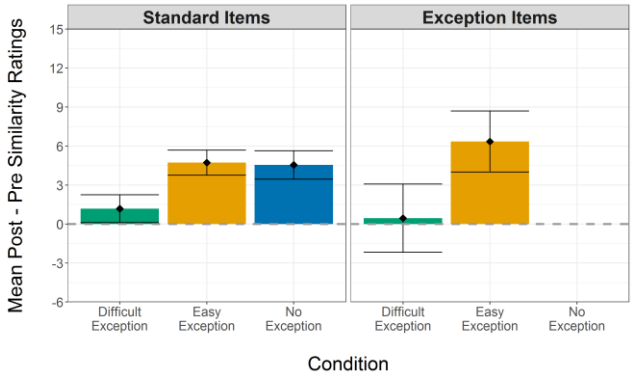

Figure 7: Between-category mean difference ratings for Experiment 2. Error bars reflect 95% CIs, diamonds reflect adjusted means obtained from the respective model. The dashed line represents no change in similarity ratings.

These results replicate the pattern of findings from Experiment 1 and suggest that the between-category compression observed in the Easy and No Exception condition are not directly related to the changes made to the number of blocks, exception frequency, or exception appearance. As discussed in Experiment 1, the best explanation pertains to be the diagonal structure that was used. Participants who do well in classification are most likely learning the diagonal slope common to both classes. Doing so highlights a similarity in the second rating phase absent in the first. In support of this conjecture, we find that the only condition which did not evidence between-category compression for standard item—the Difficult condition—also resulted in the lowest accuracy during learning of standard items.

General Discussion

The aims of the present work were to determine the extent to which exception items affect compression and expansion. More specifically, we predicted that learned exceptions would alter the ostensible cohesiveness of the category structure, resulting in reduced or mitigated compression. We further predicted that this would apply to exceptions with high similarity to the contrast category, and to a lesser degree for exceptions dissimilar to both categories. Both kinds of exceptions violate category norms, but only the former might require building a discrete representation for good performance. To the extent that the exception representation is individually weighted, it would affect perceived category cohesiveness to a greater extent. Though we also check for presence/absence of CP effects, our primary prediction was that there would be comparatively less compression for the target category in the Difficult condition compared to the Easy and No Exception condition. Experiments 1 and 2 provide mixed support for these predictions.

In Experiment 1, compression was evidenced for Category A in all conditions. Contrary to initial predictions, the Difficult condition did not impede compression for the category with the exception, however, it should be noted that participants did not learn the difficult exception either. Notably, Category B of the Difficult condition did not evidence compression. Though inconclusive, this may suggest that the presence of the nearby exception, in conjunction with corrective feedback indicating that it was not a B item, was sufficient to alter perceived cohesiveness for that category. Despite this, we were unable to find the relative differences we had predicted.

Experiment 2 provides more direct evidence of the hypothesized relationship. Participants were able to learn that the difficult exception belonged to Category A. Further, Category A in the Difficult condition did not evidence compression. The relative differences in compression between categories suggests that the mere presence of an exception (regardless of initial similarity) may be sufficient to mitigate compression effects among the standard items. It is likely that the Easy condition produced less compression compared to the No Exception condition in Experiment 2, but not Experiment 1, due to the greater number of learning trials in Experiment 2. The effect appears to be driven by greater compression in the standard items for the No Exception condition, rather than less compression for the standard items in the Easy condition.

The results from Experiment 2 suggest that with enough exposure, any exception, regardless of similarity to a contrast category, can impede the effects of category compression for the standard items that share its category. The finding that compression did not differ between the Easy and Difficult condition for the target category was unexpected. Contrary to our initial predictions, it may be that the mere presence of an exception is sufficient to impede compression, and that there is no further relationship between the difficulty of exception acquisition and the amount of compression observed. Further work is needed to clarify this relationship, or lack thereof.

Contrary to typical findings in the learned CP literature, we also found robust evidence for an effect of between-category compression. It could be argued that a lack of expansion is owed to participants not learning the categories well enough, but this account does not explain the compression observed, which we would not expect to occur in the absence of learning. Instead, we conjecture that this novel, albeit unexpected, finding may be due to the diagonal structure used. Successful performance on this structure requires learning feature correlations that are common to both classes. Learning that feature correlation may cause it to appear more salient in the second rating phase, consequently increasing the similarity between two items from opposing categories.

We briefly note that the only other study we are aware of that employs a diagonal category structure in the study of CP effects—Pothos and Reppa (2014)—did not find any evidence of either between-category compression or expansion. That said, there are an appreciable number of differences in design and materials between our study and theirs, any number of which could independently or jointly moderate the results.

There are a few limitations to the current study that should be addressed in future work. We have assumed that participants are forming distinct representations for the difficult exception and not the easy exception. We have also assumed that unique representations disrupt perceived category cohesiveness. Implementing a separate recognition phase (typically used to index exemplar distinctiveness in exceptions studies) and querying category cohesiveness after every block would provide measures that could ameliorate these concerns. Future work should also more closely investigate the nature of between-category compression—whether it is idiosyncratic to our design and materials, and what effects it might impart for higher order reasoning. To ensure that our general findings are robust across different measures of CP, efforts should be made to replicate across a broad range of indices such as a change-detection or an XAB task.

It is often speculated that learned CP effects occur to facilitate classification (Livingston et al., 1998) by making it easier to assign items to classes. The current work presents an important caveat with regards to compression—namely, that compression relies not only on task-pressures, but the ease with which a category can be compressed, as determined by category cohesiveness.

References

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814-823.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27-43.

Kurtz, K. J. (1996). Category-based similarity. In Garrison W. Cottrell (ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 290). La Jolla, CA.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732.

Pothos, E. M., & Reppa, I. (2014). The fickle nature of similarity change as a result of categorization. *Quarterly Journal of Experimental Psychology*, *67*(12), 2425-2438.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*(4), 534.

Savic, O., & Sloutsky, V. M. (2017). Peculiarity doesn't trump ordinarity: On recognition memory for exceptions to the category rule. In G. Gunzelmann, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 290). Austin TX: Cognitive Science Society.