Title

Not All Exceptions Are the Same: Different Memory Demands for Differentiation, Isolation and Odd-ball Exceptions

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Publication Date

2019

Peer reviewed

Not All Exceptions Are the Same: Different Memory Demands for Differentiation, Isolation and Odd-ball Exceptions

Abstract

There is an influential body of research arguing that category exceptions have a special status in memory compared to regular category members. However, the memory advantage for category exceptions has typically been demonstrated using one very specific category structure (Differentiation). Here we present a study examining whether the reported memory advantage is specific to this particular structure or whether it can be generalized to other kinds of exceptions (Isolation and Odd-ball). We compare three different types of category exceptions that have varying memory demands due to different levels of feature binding required for accurate categorization. The results suggest that only those exceptions that require binding together multiple features are remembered better than regular, rule-following items. The present work clarifies that the memory advantage for exceptions characterizes certain kinds of exceptions rather than exceptions in general.

Keywords: category exceptions; rule-plus-exception; binding requirement

Introduction

Whales are mammals. Penguins are birds. Tomatoes are fruit. Many categories include items that look different, behave differently, or lack important qualities that define the majority of members of the category. We refer to these items as *exceptions,* because they violate our expectations about the category.

Since the goal of categorization is to encode key aspects about the members of the category, it is reasonable to ask: How are exceptions, items that violate those key aspects, learned and represented?

Memory Advantage for Exceptions

There is an influential body of research arguing that category exceptions have a special status in memory. Palmeri and Nosofsky (1995) demonstrated that exceptions to a category rule are remembered better than the items that follow that rule. This initial finding of a memory advantage for category exceptions is supported by a number of subsequent category learning studies (Sakamoto & Love, 2004, 2006; Davis, Love & Preston, 2012) and found to be in accordance with the predictions of several influential models of category learning: RULEX (Nosofsky, Palmeri & McKinley, 1994) and SUSTAIN (Love, Medin, & Gureckis, 2004).

Work on this topic in the categorization literature was preceded by studies in memory (Von Restorff, 1933) and schema research (e.g. Rojahn & Pettigrew, 1992; Stangor & McMillan, 1992), where an advantage in memory for exception items has long been established. Although the approach (both in terms of methodology and primary research questions) differed between the memory and categorization literatures, the fact that these findings paralleled each other further strengthen the view that there is a general advantage in memory for information that does not fit in with salient knowledge structures.

What Makes (Some) Category Exceptions Memorable?

Although a memory advantage for category exceptions seems to be well established, the nature of the effect is not well understood. One obstacle to understanding what makes category exceptions more memorable is that previous studies focused on one very specific type of category structure. We refer here to this structure as the Differentiation case (see Figure 1).

Although exceptions are not limited to the Differentiation case, the vast majority of influential work on this topic (Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004; Davis, et al., 2012) has studied this kind of exception. One reason the Differentiation structure has received so much attention is that it presented an interesting challenge for models of category learning to explain. And so, researchers often selected structures with the purpose of evaluating or comparing models of categorization, and were not necessarily concerned with representing all types of exceptions.

Thus, it remains unclear whether the reported memory advantage for category exceptions is specific to this particular structure or whether it can be generalized to other kinds of structures. In what follows, we describe how the difference in category structures may affect memory demands.

Different Memory Demands for Differentiation, Isolation and Odd-ball Exceptions

Figure 1 illustrates three different structures of exception items: Differentiation, Isolation and Odd-ball exceptions. All three types of exceptions (a) violate the category rule and (b) are dissimilar to other items in their own category. However, they differ in how much they share with the contrasting category members.

Differentiation exceptions, the most commonly used structure, are highly similar to items of the contrasting category. Not only do they follow the contrasting category rule, but they also share other features with items of the contrasting category. Due to its specific structure, this kind of exception cannot be categorized correctly on the basis of any one individual feature.

Since it is not enough to remember one or even multiple isolated features, Differentiation exception features have to be bound together and committed to memory. This is because each feature of the exception (in our case, the color, the size and the shape) has a competitor in the contrasting category, thus making only the whole configuration (but not individual features) sufficient. For example, the past tense of the irregular verb teach (taught) is very different from that of similar sounding regular verbs (reach, breach, or preach), but is similar to that of a phonetically different verb (e.g., think).

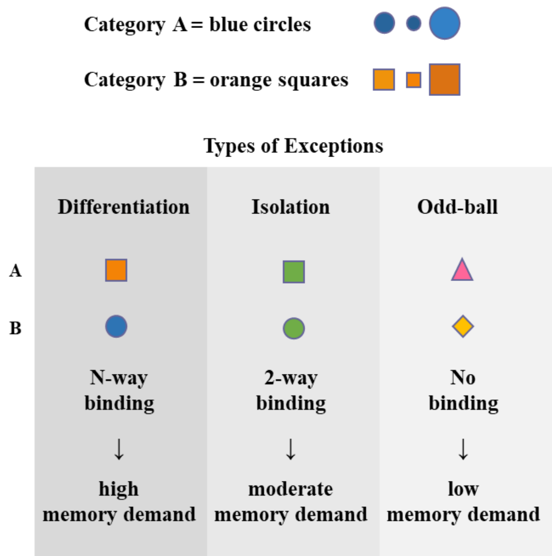

Figure 1: Three different types of exception items for categories of blue circles and orange squares. Exceptions vary in how similar they are to the contrasting category and therefore in the amount of feature binding and memory demands required for accurate categorization.

Although the Differentiation structure is commonly used in experimental studies, this kind of exception is likely quite rare in the real world. Most exceptions found outside of the lab, even the most commonly cited examples (e.g. bats as exceptions to category of mammals), in addition to shared features also have some distinctive features. Figure 1 illustrates two types of such exceptions: Isolation and Odd-ball.

Isolation exceptions follow the contrasting category rule, just like Differentiation exceptions, but, crucially, they also have distinctive features. The distinctive features make Isolation exceptions less similar to items of the contrasting category. As fewer features are shared, less complex binding is required, which reduces memory demands. In the example shown in Figure 1, the green square has the same shape (rule dimension) as members of the contrasting category. However, its unique color (different from both its own and the contrasting category) allows for the categorization problem to be solved based on binding of only two dimensions: color and shape. Analogous to this example, bats are flying creatures (characteristic of the contrasting category of birds) with membranous wings (a distinctive feature), and therefore could be represented as exceptional mammals by binding these two features: flying and membrane wings.

The third kind of exceptions shown in Figure 1 is the Oddball exception. Odd-ball exceptions do not share any features with members of contrasting category. Since all of their features are distinctive, no binding is required, and categorization can be made on the basis of any single feature. Critically, in case of the Odd-ball structure, representation of exceptions can be as simple as representation of regular items. The pink triangle in Figure 1 violates both shape and color of the category of blue circles. However, since there is no overlap on these dimensions with any items in category B, accurate categorization can be based on either its pink color or triangular shape alone. One example of such an oddball is an hourglass as an exceptional member of the category of time-keeping devices.

Present Experiments

Although all three of the different structures presented in Figure 1 represent rule-violating exceptions, they have different memory demands. Since the categorization literature has focused almost exclusively on the Differentiation case, it remains unclear whether in previous studies exceptions were remembered better (a) because they violated a salient knowledge structure (von Restorff, 1933; Hunt & Lamb, 2001; Busey & Tunnicliff, 1999; Nairne, 2006), or (b) because of the additional binding requirement resulting from the high overlap with the contrasting category (Sakamoto & Love, 2006).

In support for the latter possibility, Sakamoto and Love (2006) demonstrated that exceptions that are more similar to the contrasting category (i.e., Differentiation) are remembered better than exceptions that are more distinctive (i.e., Isolation). The stimuli in Sakamoto and Love (2006) were lines that varied in color and size and the focus of this study was on comparing different exceptions to each other. Thus, memory advantage for both kinds of exceptions over regular items was assumed although not directly tested due to the limitations of the stimuli design.

The present experiments were designed to tease apart the roles of (a) violation of a salient knowledge structure (i.e. similarity of an exception to its own category) and (b) differences in binding requirement (i.e. similarity of an exception to the contrasting category).

Three experiments were conducted. Experiment 1 was set as a replication of previous studies that examined recognition memory for Differentiation structure. The category structure completely follows the one reported by Davis, Love, & Preston (2012) and uses the same experimental tasks and procedures. Experiment 2 and Experiment 3 build on Experiment 1 by employing the same experimental design, procedures and materials to examine memory for Isolation (Experiment 2) and Odd-ball (Experiment 3) exceptions.

If the memory advantage for category exceptions results from violation of a salient knowledge structure, exceptions should be remembered better than regular items across the three experiments. However, if the advantage for exceptions is dependent on differences in binding requirements, Differentiation (and potentially Isolation exceptions) should be remembered better, while there should be no advantage for Odd-ball exceptions.

Experiment 1: Differentiation

Methods

Participants Participants were 38 undergraduate students from a Midwestern university who received course credit for their participation. Two additional participants were excluded due to the failure to finish the experiment.

Materials The stimuli were schematic clown-like faces that varied along four (feature) dimensions. Items were accompanied by two novel category labels: Zuzu and Tati. As it can be seen in Figure 2, the four feature dimensions were hair, eyes, mouth and side whiskers. Side whiskers were selected as a rule dimension and the three other features varied between the two categories.

The category structure Table 1 shows an abstract representation of the category structure used in Experiment 1 (as well as ones used in Experiments 2 and 3). Each of the two categories had three Regular, rule-following items and one Exception.

Rule-following items could be categorized accurately based on the value of a single rule-dimension (side whiskers). The rule dimension was held constant across participants. The other three dimension varied between the categories, with exactly the same combinations of the three features used for constructing items of category A and category B (see Table 1). The Exceptions appeared to belong to the opposing category based on their value on the rule-relevant dimension. Additionally, the two exceptions had the same values on the three other dimensions, and thus could be categorized accurately only based on the representation that captures the combination of the rule and (at least) 2 other features.

Based on the items presented in Table 1 that were used during training, we constructed foils for memory test. The foils had the same feature values as training items, but in novel combinations. There was a total of 8 foils constructed for memory test in Experiment 1.

Table 1: The category structure used in Experiments 1-3

	Category A	Category B
Regular items		
same set across the three experiments	**1** 334	**2** 334
	1 343	**2** 343
	1 433	**2** 433
Exceptions		
Exp 1: Differentiation	**2** 444	**1** 444
Exp 2: Isolation	**2** 555	**1** 555
Exp 3: Odd-ball	**8** 888	**9** 999

Note. 1 = rule of category A; 2 = rule of category B; 3 = probabilistic; 4 = probabilistic; 5 = novel non-diagnostic; 8 = unique for exception A; 9 = unique for exception B.

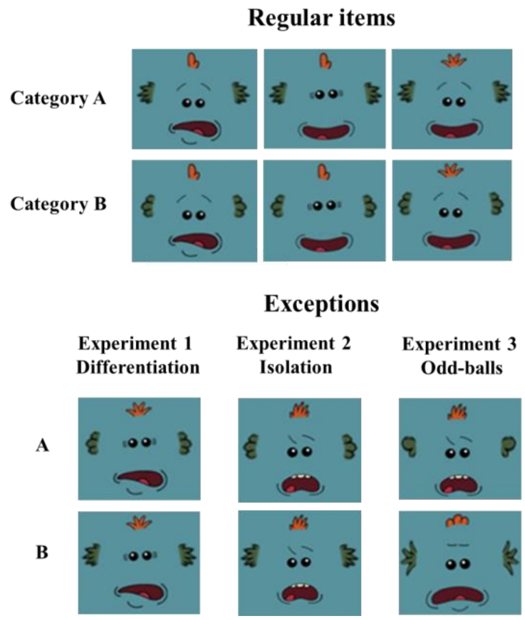

Figure 2: Complete set of stimuli used in the training (Experiments 1–3). Regular items were the same across the experiments.

Procedure

The experiments consisted of three phases: training, memory test and categorization test.

Training During training participants were presented with the exemplars of the two categories and were asked to classify each exemplar.

Following the procedure of previous studies, participants were given explicit instructions indicating the rule feature. They were encouraged to use this feature during categorization and to memorize items that violate this rule (Davis, Love, & Preston, 2012).

Items were presented individually, and corrective feedback was provided after each response. There was a total of 64 trials presented during training, 48 rule-following items and 16 exceptions (i.e. each of the 6 Regular and 2 Exception items presented 8 times in random order).

Memory test Following training, participants were introduced to the memory test. In the memory test, participants saw two items at a time: one training item and one foil (item that had the same features as the training items, but in a novel combination). Their task was to say which of the two items was old (presented during the training). There was no feedback given during memory test. The test had 48 trials presented in a random order.

Categorization test In the categorization test participants were presented with Regular and Exception items they saw during the training. The procedure was exactly the same as in the training session with the only difference being that during the categorization test, participants were not provided with feedback. There were 16 categorization test trials, 8 Regular items and 8 Exceptions.

Results

Figure 3 shows participants' recognition memory and categorization accuracy (panel a).

Participants were less accurate at categorizing Exceptions ($M = 0.49$, $SD = 0.37$) than Regular items ($M = 0.82$, $SD = 0.24$), $t(37) = 5.06$, $p < 0.001$, $d = 0.82$. However, they had better recognition memory for Exception items ($M = 0.63$, $SD = 0.19$) than for Regular, rule-following items ($M = 0.47$, $SD = 0.12$), $t(37) = 3.77$, $p < 0.01$, $d = 0.61$.

Both of these results, memory advantage for Exceptions and better categorization accuracy for Regulars, are in accordance with previously reported findings (Palmeri & Nosofsky 1995; Sakamoto & Love, 2004, 2006).

It is important to note that here, as in the previously reported studies, the advantage in memory for rule-violating exceptions results from optimization in memory for Regular items. Since participants categorized Regular items relying on the category rule, the rule feature is the only feature they needed to represent and thus they had no need to remember individual exemplars representing the category. On the other hand, in order to learn Exceptions, they had to bind in memory information about a minimum of three features (the rule and two other features).

Experiment 2: Isolation

The goal of Experiment 2 was to test the robustness of the memory advantage for Exceptions, when rule-violating items are Isolation Exceptions. Isolation Exceptions have lower memory demands than Differentiation Exceptions, but they still require binding of information about the rule and (at least) one more feature.

Methods were identical to Experiment 1, except for the type of Exception participants needed to learn (See Table 1). Foil items for the memory test were designed accordingly to include features of Isolation Exceptions, which resulted in 19 foil items in total. Twenty-three undergraduates from a Midwestern university participated for course credit.

Results

The pattern of results in Experiment 2 closely replicated the one observed in Experiment 1 (Figure 3).

Although participants were more accurate when categorizing Regular ($M = 0.85$, $SD = 0.18$) than Exception items ($M = 0.65$, $SD = 0.36$), $t(22) = 2.41$, $p < 0.05$, $d = 0.50$, they remembered Exception items ($M = 0.67$, $SD = 0.26$) more accurately than Regular, rule-following items ($M = 0.51$, $SD = 0.09$) , $t(22) = 2.91$, $p < 0.01$, $d = 0.61$.

Experiment 3: Odd-ball

The critical difference between Experiment 3 and Experiments 1-2, was that learning of rule-violating items in Experiment 3 did not require forming of a complex binding structure. Both Regular and Exception items could be categorized based on a single, individual feature. Thus, any differences in recognition memory between Regular items and Odd-ball Exceptions could be solely due to effects of rule violation.

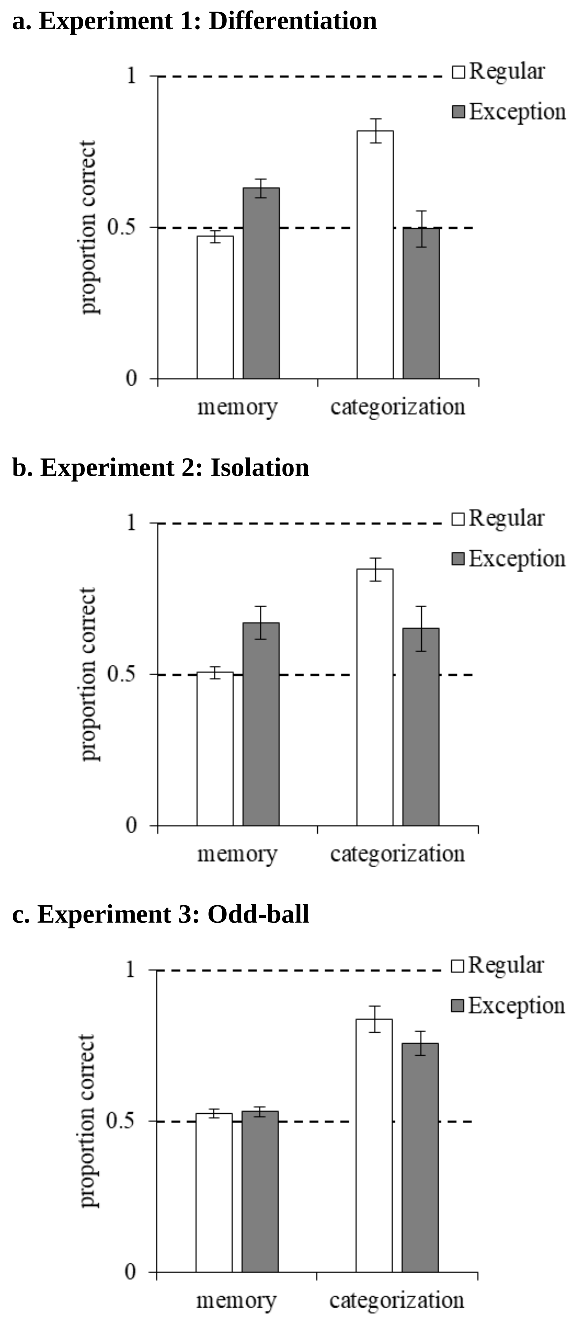

Figure 3: Recognition memory and categorization accuracy across the three experiments. Error bars represent standard errors of mean.

Methods were identical to the ones used in Experiment 1-2, except that Experiment 3 used Odd-ball Exceptions as the rule-violating items (See Table 1). Foil items for the memory test were constructed following the logic of Experiment 1 and 2. There was a total of 21 foil items used for the memory test. Forty undergraduate students from a Midwestern university participated for course credit.

Results

In Experiment 3, participants were equally accurate when categorizing Exceptions (M = 0.76, SD = 0.25) and Regular items (M = 0.84, SD = 0.28), $t(39) = -1.30$, $p = .203$, d = 0.20.

Critically, we observed no difference in recognition memory between the two item types, $t(39) = 0.18$, $p = .858$, $d = 0.03$. Participant had no memory for exemplars of either Regular (M = 0.53, SD = 0.10), or Exception items (M = 0.53, SD = 0.11). One sample t-tests against chance were approaching significance for both Regulars ($t(39) = 1.79$, p = .081) and Exceptions ($t(39) = 1.78$, $p = .084$).

Discussion

The presented work aimed to clarify whether category exceptions merit a special memory representation because they violate a salient knowledge structure, as it has been previously assumed (e.g., von Restorff, 1933; Hunt & Lamb, 2001; Busey & Tunnicliff, 1999; Nairne, 2006), or is it only those exceptions that have high binding requirements that have the special memory status. This is a critical question, as in the latter case, the special memory status characterizes certain kinds of exceptions rather than exceptions in general. Consequently, the generalizations often present in the categorization literature when discussing exceptions would be unjustified.

The recognition memory comparisons across the three experiments revealed that participants had better recognition memory for exceptions that required binding of two or more features to be accurately categorized. However, when memory demands for regular and exception items were equal, there was no memory advantage for exceptions. In other words, when category structure does not require feature binding for successful categorization and both item types can be classified based on the individual features, exception items are treated as any other regular item. Participants may optimize their memory when learning exceptions in the same manner as they do when they learn regular items, and thus have poor exemplar memory for both regulars and exceptions.

Model Predictions

The results of previous studies on recognition memory for categories with exceptions were found to generally conform to the predictions of RULEX and SUSTAIN (Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004, 2006). Pure exemplar storage models, such as the context model (Medin & Schaffer, 1978), have also been considered and found to be inadequate at simultaneously predicting categorization accuracy and recognition memory for categories with exceptions (Palmeri & Nosofsky, 1995). Our results replicate this failure of exemplar models. Exemplar models have difficultly predicting good categorization, but bad memory, for regulars since categorization relies directly on memory storage. Similarly, they struggle with good memory, but poor categorization, of exceptions. In general, exemplar models would tend to predict that both categorization and memory would be better (or possibly both worse) for exceptions than regulars, but they would not predict opposite patterns for categorization and memory.

RULEX provides good predictions for the patterns that were found in the Differentiation and Isolation structures, correctly predicting better categorization of rule-following items, but poorer memory for those items since they are not represented independently in memory. Memory is predicted to be better for exceptions since they need to be stored individually in memory. It is unclear, though, what RULEX would predict for the Oddball structure.

Versions of RULEX have been formulated for binary-valued discrete dimensions, and continuous-valued dimensions (Nosofsky & Palmeri, 1998), but (to our knowledge) not for discrete dimensions with more than two possible values. While there are straightforward ways in which to extend RULEX to accommodate this type of stimulus structure, there are several alternative formulations that would provide opposite predictions.

Existing versions of RULEX first try simple unidimensional rules. If perfect rules fail, it then attempts unidimensional rules with exceptions stored in memory. If those representations are inadequate, it moves to considering more complex rules. Our Oddball category structure could theoretically be solved with a disjunctive rule on a single dimension (i.e. value 1 *or* 8 on dimension one is Category A, 2 *or* 9 is category B; see Table 1). It is unclear whether RULEX would attempt to use this type of rule prior to or after it attempts to store exceptions in memory (considering that the rule is technically unidimensional but also somewhat complex). If it tries storing exemplars first, it would predict similar behavior as in the Isolation structure (and therefore, fail to predict our results). If it tries the disjunctive rule first, then it would not need to store any exemplars in memory, and could match participants' data well. So, RULEX could predict all of our results in theory, but it depends on exactly how it is formulated to handle this type of stimulus representation.

SUSTAIN (Love, Medin, & Gureckis, 2004), on the other hand, can naturally process the stimulus structures used in our study without needing modification, but its predictions are somewhat less intuitive. Like RULEX, the predictions depend on whether it stores exceptions separately (by creating additional clusters) or together with regulars. In theory it can do either depending on the exact parameter settings and the order in which it encounters the exemplars. To test whether SUSTAIN would create different numbers of clusters for the three different stimulus structure, we performed simulations of the model. We first fit the model to each participants' training data using maximum-likelihood estimation in order to obtain reasonable parameter estimates. Then we simulated the model using each parameter combination on all three stimulus structures (1000 simulations per parameter combination, per structure).

Results of the simulations generally match the behavioral results: better categorization for regulars than exceptions in all structures, but better memory for exceptions than regulars—except in the Oddball structure where memory performance was roughly equivalent between regulars and exceptions. Additionally, the number of clusters formed was found to be highest for the Differentiation structure (median: 6 clusters; mode: 4 clusters), slightly lower for the Isolation structure (median: 4 clusters; mode: 4 clusters) and lowest for the Oddball structure (median: 3 clusters; mode: 3 clusters). Importantly, that there were typically fewer than 4 clusters in the Oddball structure indicates that exceptions were not represented completely independently of the regulars, which is consistent with worse memory for exceptions compared to the other two structures.

In summary, both RULEX and SUSTAIN can potentially account for our results by representing exceptions separately from regulars in the Differentiation and Isolation structures, but not in the Oddball condition. SUSTAIN produces this pattern as a normal result of its category learning process, while RULEX produces this result under one of several possible instantiations of its decision process. In both models the separate representations of exceptions are consistent with increased feature binding for those items compared to regulars, though they may not have been described in terms of feature binding in previous work.

Conclusions

Taken together, our findings suggest that the previously reported advantage for memory exceptions reflects elevated memory demands of specific kind of exception which does not generalize to other kinds of exceptions.

This work further adds to our understanding of what makes *some* category exceptions more memorable, by focusing on the critical role of competition between the exception and contrasting category members.

Acknowledgments

This information will be added to the final version of this manuscript (in order not to violate double-blind review process).

References

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 1210-1235.

Davis, T., Love, B.C., & Preston, A.R. (2012). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, 22, 260-273.

Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 1359-1366.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*, 111, 309–332.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review,* 85, 207-238.

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory*, New York: Oxford University Press.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53–79.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, 5, 345-369.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 548–568.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, 31(2), 81–109.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, 133, 534–553.

Sakamoto, Y., & Love, B.C. (2006). Vancouver, Toronto, Montreal, Austin: enhanced oddball memory through differentiation, not isolation. *Psychonomic Bulletin & Review*, 13, 474-9.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and developmental literatures. *Psychological Bulletin*, 111, 42–61.

von Restorff, H. (1933). Analyse von vorgängen im spurenfeld: I. Über die wirkung von bereichsbildungen im spurenfeld [Analysis of processes in the memory trace: I. On the effect of group formations on the memory trace]. *Psychologische Forschung*, 18, 299–342.