Title
Category Exceptions Change Category Boundaries

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Savic, Olivera
Sloutsky, Vladimir

Publication Date
2022

Peer reviewed

Category Exceptions Change Category Boundaries

Olivera Savic (savic.7@osu.edu)
Department of Psychology, 1835 Neil Avenue,
Columbus, OH 43210, United States

Vladimir Sloutsky (sloutsky.1@osu.edu)
Department of Psychology, 1835 Neil Avenue,
Columbus, OH 43210, United States

Abstract

In order to successfully guide generalization of knowledge, category representation needs to be both: flexible enough to account for new evidence and stable enough to resist harmful change. Here we present a set of experiments designed to test how items that violate our expectations (i.e., category exceptions) affect category representation. Specifically, we wanted to know whether learning a category exception can change category boundaries. Does learning about penguins changes the way we think about birds? Do features of penguins contribute to making decisions as to whether a novel item is a bird? Across two experiments we found evidence that exceptions can change category boundaries and thus significantly affect future generalization. We discuss implications these findings have for the extent models of category learning and memory.

Keywords: category learning; generalization; exceptions; category boundaries

Introduction

One of the key hallmarks of human intelligence is the ability to extend our knowledge from familiar to novel. Based on our experience with fruit, we can infer that quince is edible even if have never encountered this fruit before (Gelman, 2009; Gelman & Meyer, 2011; Sloutsky, 2010). Therefore, our ability to generalize knowledge allows us to use what we have learned in the past, to predict future. As such, this ability represents a key tool humans need in order to successfully deal with the uncertainty of the everyday experience.

Although it is typically performed in a fast and effortless way, even the simplest forms of generalization may be prone to error (Heit, 2000; Rips, 2001; Rehder, 2006). Most importantly, to make a generalization our cognitive system needs to decide which aspects of the previous experience are useful for making predictions about the novel item. For example, how useful is our experience with penguins in making predictions about the other birds?

The current study aims to broaden our understanding of the principles that guide generalization by examining the effects that deviant items, such as flightless birds, have on category based generalization in adult human learners.

Category based generalization: Challenges

If it has a beak, wings and feathers, it lays eggs and flies, it is a bird. This understanding of what makes a bird is likely based on our experience of pigeons, crows, robins, cranes, and other flying birds. It is useful and accurate. However, it does not account for all members of this category. For example, penguins are birds although they do not fly.

Our everyday experience is filled with examples such as this one, where a new experience violates the expectations we formed based on the salient regularities in the environment. There are different solutions our cognitive system may use to solve this problem. One possibility is that the exception and regularity-following items are represented independently. In this case, no change in the previously formed representation of the category is required. In language this is needed in order to avoid overgeneralizations (e.g., Yang, 2016). For example, when learning English one must learn when to generalize a property from familiar to novel verbs (e.g., add "–ed" to form past tense), and when properties apply only to specific verbs (e.g., eat – ate). Similar idea, the need for independent representation of exceptions, was suggested in categorization literature in order to explain advantage in memory for exceptions over regular category items (e.g., Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004).

On the other hand, our cognitive system may try to integrate the information about exception with knowledge about the regularity-following items. Based on this view, learning about penguins could change the representation of birds, so that it includes not only features of typical and common birds, but also the features characteristic of penguins. While both of these solutions could result in equally successful classification of penguins, they would support different category decision about novel birds we may encounter, such as emus or kiwis. If we follow the first solution, emus and kiwis would not be classified as birds since they lack the key property - flying. On the other hand, if penguins expand category boundaries, kiwis and emus would be classified as birds, without a need to learn this information anew.

Importantly, both of the described solutions come at a cost. If we keep the representation of birds as flying creatures, we risk to misclassify rare members of the category. Alternatively, if we update the representation of birds to include features of penguins, we risk to include irrelevant features in category representation. Category representation that includes both relevant and irrelevant features results in erroneous generalizations driven by irrelevant features (Sloutsky, 2010; Deng & Sloutsky, 2016; Castro, Savic,

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Navarro, Sloutsky & Wasserman, 2020). How does human cognitive system resolve this challenge?

Although category-based generalization was extensively tested in categorization literature (e.g., Johansen & Palmeri, 2002; Erickson & Kruschke, 2002; Lacroix, Giguere & Larochelle, 2005), we know little about the effects of *exceptions* on generalization. However, work on category exceptions provides important insights about how exception items may be represented. As hinted in the previous section, one of the dominant views in the literature, and the one that inspires this work, has been that exception and regular items need to be stored independently. It is important to understand that this solution was suggested by some influential models of category learning (e.g., RULEX, Nosofsky, Palmeri, & McKinley, 1994; SUSTAIN, Love, Medin & Gureckis, 2004) in response to resolving discrepancy in learning and memory for rule-violating exceptions – while regular items tend to be learned more successfully, exceptions tend to be remembered better. Importantly, these models have not been used to predict the effects of exceptions on generalization. Therefore, we will only discuss the importance of the theoretical assumption proposed by these models which shaped the dominant view in the literature - the view that exceptions need to be represented separately from regular items. One of the key implications of this assumption is that the generalization based on the categories with exceptions will be mainly shaped by the regularities that define the regular category members. Therefore, features of exceptions (i.e., penguins) will have small (or no) contribution in making category decisions compared to the features of the regular items.

Combining insight from empirical studies and modeling, Erickson and Kruschke (2002) demonstrated that participants tend to classify novel stimuli according to the rule followed by the majority of the category members (regular items), even when the new item is most similar to the exception. This work seems to suggest that the regularity that defines majority of the category members may not only play an important role in generalization, but may also override the contribution of the features of the exception items.

Therefore, in addition to being an interesting empirical question, understanding what effect category exceptions have on generalization has important theoretical implications.

Present Study

The present study was designed to test the effect of category exceptions on generalization. Specifically, we examined whether and how categorization of novel items changes after participants learn exceptions.

Experiment had two parts. Each part had a training and a testing phase. In the first part, participants were trained to assign regular items to two categories. In the second part of the experiment, they learned that one of these categories has an exception. To examine the effects of exceptions on generalization, participants were tested on the same set of novel items both in the first and the second part of the experiment. Generalization items were built by combining features of the regular items and ones of the exception. Critically, these features were in conflict – while the feature of the regulars signaled membership in category A, the feature of the exception signaled membership in category B. Thus, category decisions on these items uncovered whether during generalization participants relied on the features of the exception item, or only features of the regulars.

If exceptions do not influence generalization, classification of Generalization items should not change after participants learn about the exception. In other words, participants are expected to rely on the features of the regular items throughout the experiment. However, if exceptions do affect categorization, we expect to observe a change in generalization patterns after exception has been introduced.

We should emphasize that the category structures typically used in the prior work do not allow for a direct test of whether the features of exceptions are used in generalization. This stems from the fact that exceptions are typically considered as items that have the same features as the regular items (i.e., the regular items of the contrasting category). For example, for categories of pink circles and yellow triangles, an exception to the category of pink circles would be a yellow triangle. While this type of exceptions is useful for testing effects of exceptions on memory, it is clearly not well suited for evaluating the unique contributions of features of exceptions to generalization of new items. Therefore, in the current study we designed the exceptions as items that violate salient regularities, but have unique features.

Experiment 1

Methods

Participants We recruited 29 undergraduate students from a large Midwestern university. They received course credit for their participation.

Stimuli Stimuli were two-dimensional items that varied in shape and color (see Figure 1).

Category structure As illustrated in Table 1, category items had one rule feature that perfectly defined category membership of the regular items and one feature that was highly predictive (3 out of 4 regular items had this feature).

There were three types of stimuli items: Regular items, Exception and Generalization items.

Regular items belonged to two categories: A or B. Each category had four items: 3 Prototypes and 1 High Match item. Category A was a category of pink items, 3 pink circles and one pink triangle. Category B was a category of yellow items, three yellow triangles and one yellow circle (see Figure 1). Therefore, category membership was fully predictive based on the color, the rule dimension. The other dimension, shape, was probabilistic. One out of four training items in each category had the contrasting category value on this dimension (see Table 1). Items did not vary on other dimensions but color and shape. Therefore, although each category had 4 items, 3 of these items (i.e., Prototypes) were identical.

Exception item was different from members of category A and category B on both of the dimensions, color and shape.

For the above described regular items, Exception was a turquoise square (Figure 1). Please note that Exception was an individual. Therefore, both of its features were fully predictive. However, we will refer to the dimensions of the Exception having in mind the overall category structure which differentiates between the rule and the probabilistic feature.

Generalization items were hybrid items that combined the features of Regular items and the Exception. There were two types of these items. The first had the rule feature of the Regulars and the probabilistic feature of the Exception. The second had the opposite combination: the rule of the Exception and the probabilistic feature of the Regulars. Critically, the two features of Generalization items were in conflict, signaling different category membership. Figure 1 illustrates examples of generalization items when Exception (turquoise square) was a member of the category A (pink items). In this example, yellow square has the Rule of the category B Regulars (yellow items) and the probabilistic feature of category A Exception (square shape). On the other hand, turquoise triangle has the Rule of the category A Exception (turquoise color), and the probabilistic feature of category B Regular items (triangle shape).

Table 1: Category structure used in Experiments 1-2. The rule dimension is in bold font.

Training items				
	Category A		**Category B**	
Regular (Prototype)	**0**	0	**1**	1
Regular (Prototype)	**0**	0	**1**	1
Regular (Prototype)	**0**	0	**1**	1
Regular (High Match)	**0**	1	**1**	0
Exception	**2**	2		
Test Items				
T1	**1**	2		
T2	**2**	1		

Design All participants were trained and tested in both Baseline and Exceptions condition (Condition; within-subjects; levels: Baseline, Exception). However, for one half of the participants Exception was assigned to the category A (N=14) and for the other half (N=15) to the category B (Exception Assignment Condition; between-subjects; levels: Category A, Category B). All participants were tested on the same types of generalization items.

Procedure The experiment had four parts: Baseline Training, Baseline Test, Exception Training and Exception Test.

The task was the same at Training and at Test. Participants were presented with an item (e.g. a pink circle) and asked to classify it in one of the two categories (see Figure 2). The only difference between Training and Test was that at Training participants received corrective feedback after they gave a response (e.g. “That’s right! It is for Sony.”), while at Test there was no feedback.

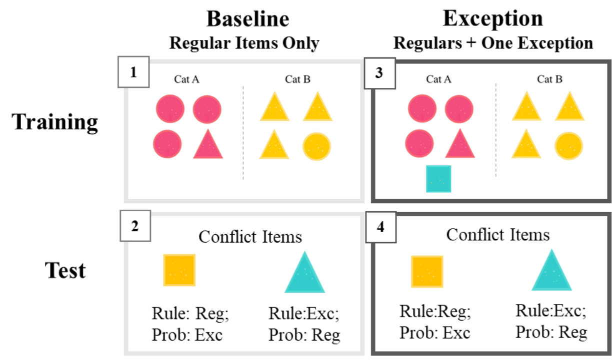

Figure 1: Illustration of experimental design and category structures used in Experiments 1 and 2. The figure gives illustration of version of the experiment in which Exception is a member of the Category A.

In Baseline Training participants saw all of the Regular Training items twice (16 trials in total). In Exception Training, in addition to 16 Regular Training items, participants saw the Exception 4 times (20 trials in total). Baseline Test and Exception Test had identical set of items: two types of test items (Table 1) and Regular training items. Each type was presented 8 times (24 trials in total). The decision to present each test item type 8 times was made based on the previous categorization literature (Deng & Sloutsky, 2016; Savic & Sloutsky, 2019).

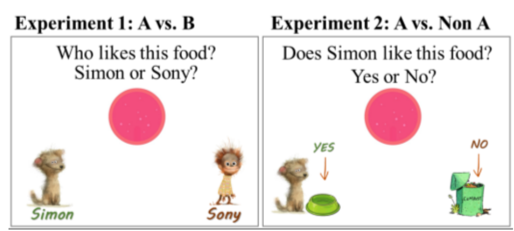

Figure 2: Illustration of the task design in Experiments 1 and 2.

Results

Training Learning was tested based on the accuracy participants achieved in the phase of the experiment in which they were first presented with an item type: Baseline phase for Regular items, and Exception phase for Exceptions.

Participants performed with an average accuracy of .92 (SD = .11) on Regular items and .78 (SD = .21) on Exceptions, which was well above the chance performance of .50 (ts > 7.35, ps < .001). There were no differences in learning of Regulars or Exceptions between the two Exception Assignment Condition groups (both ps > .10)

Test Generalization was tested based on the performance on two Generalization items (see Figure 1). To test the change in performance between the conditions, we looked at how likely participants were to classify Generalization items as members of category A (i.e., Proportion of category A responses), both before they learned to categorize the Exception item (Baseline) and after (Exception).

As it can be seen in Figure 3 (panel A), learning about the Exception did not change classification of the Generalization item that followed the rule of the Regulars and had the probabilistic feature of the Exception. On the other hand, introducing the Exception significantly changed classification of the Generalization item that followed the rule of the Exception and had the probabilistic feature of Regulars (Figure 3, panel B).

The pattern seen in Figure 3 was confirmed by two two-way mixed ANOVAs with Condition (within subjects: Baseline vs. Exception) and Exception Assignment Condition (between subjects: Category A vs. Category B) as factors and proportion of category A responses as a dependent measure.

The analyses revealed that performance on items that had the rule of the Regular item and the probabilistic of the Exception was not different before and after they learned Exception. Assignment Condition was significant ($F(1, 27) = 958.4$, $p < .001$, $\eta^2_G = .957$), but Condition and their interaction were not significant (*ps* > .10).

On the other hand, we found a significant interaction of the two factors ($F(1, 27) = 12.1$, $p < .01$, $\eta^2_G = .191$), as well as significant effect of the Assignment Condition ($F(1, 27) = 56.67$, $p < .001$, $\eta^2_G = .498$) in performance on items that had the rule feature of the Exception and the probabilistic feature of the Regulars item (turquoise circle).

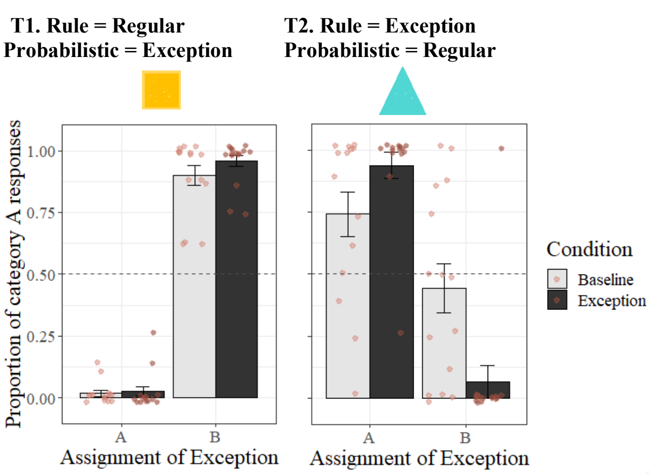

Figure 3: Performance on Generalization items in Experiment 1. Figure shows how likely participants were to classify Generalization items as members of category A, before (Condition: Baseline) and after (Condition: Exception) they learned the Exception. The performance is presented for two Exception Assignment Conditions (Category A vs. Category B). Error bars represent standard errors of mean. Pink dots represent individual participants. The dotted line marks chance performance of .50.

Experiment 2

The aim of Experiment 2 was to test whether exceptions have the same effect on generalization when participants are not forced to classify novel item in one of the two defined categories (i.e., "things can be either A or B"). We wanted to test whether the pattern reported in Experiment 1 generalizes to situations of more flexible classification where participants can also classify novel items as members of an open category (e.g., "something else"). In other words, participants' may classify an emu as a bird when they are asked to choose between a bird and a fish, but not when they can choose between "a bird" and "not a bird".

Experiment 2 used the same stimuli, task and procedure as Experiment 1, with one important difference. While in Experiment 1 we used a standard design in which items can be classified in one of the two defined categories (i.e. item either belongs to the category A or the category B), in Experiment 2, participants were instructed to classify items as either belonging to the category A, or not belonging to the category A. In other words, instead of using two defined categories (the design of Experiment 1), here we used one defined category (e.g., category A) and one open category (e.g., non A).

In Experiment 2, participants were 44 undergraduate students from a large Midwestern university who received course credit for their participation. Twenty-four of these participants took part in version A of this experiment, and 20 took part in version B.

Results

The logic and the steps in data analyses were the same as in the Experiment 1.

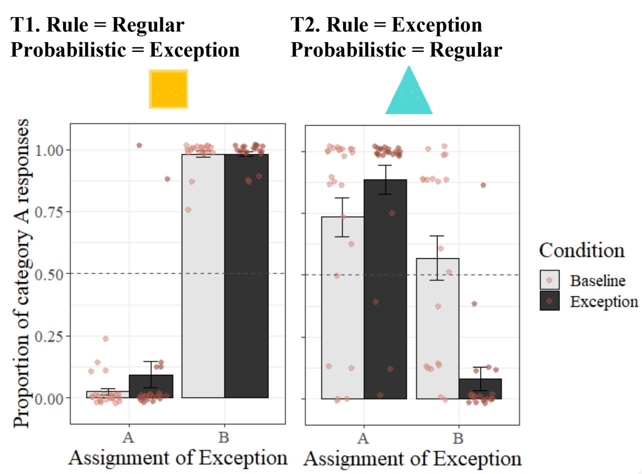

Figure 4: Performance on Generalization items in Experiment 2. Figure shows the proportion of Category A responses across two Conditions (Baseline vs. Exception) and two Exception Assignment Conditions (Category A vs. Category B). Error bars represent standard errors of mean. Pink dots represent individual participants. The dotted line marks chance performance of .50.

Training Accuracy was high on both types of training items. Participants performed with an average accuracy of .91 (*SD* = .11) on Regular items and .85 (*SD* = .21) on Exceptions, which was well above the chance performance of .50 (*ts* >

11.46, *ps* < .001). There were no differences in learning of Regulars between the two Exception Assignment groups ($p > .10$), but Exception was learned better in group in which Exception was assigned to category B, $t(35) = 2.34$, $p = .02$. However, in both groups average accuracy on Exceptions was high (.89 and .76).

Test The analyses revealed that performance on the generalization item that had the rule feature of the regular items and the probabilistic of the Exception was affected only by the Assignment Condition ($F(1, 42) = 808.5$, $p < .001$, $\eta^2_G = .911$), while Condition and their interaction were not significant (*ps* > .10).

On the other hand, we found a significant interaction of the two factors ($F(1, 42) = 21.5$, $p < .001$, $\eta^2_G = .197$), as well as a significant effect of Exception Assignment group ($F(1, 42) = 45.6$, $p < .001$, $\eta^2_G = .361$) and Condition ($F(1, 42) = 5.97$, $p < .05$, $\eta^2_G = .064$) on performance on the generalization items that had the rule feature of the Exception and the probabilistic feature of the Regular items.

The pattern of results in Experiment 2 thus completely replicates the pattern reported in Experiment 1.

Discussion

Across two experiments reported here, we have demonstrated that category exceptions affect generalization. This was true even though (a) participants were introduced to exception only after they formed a robust representation of regular items, (b) exceptions were rare – regular items were 8 times more frequent, (c) exceptions were not confusable with regulars and there was no overlap in features, and (d) participants were not forced to rely on the feature of exception – each test item could also be classified based on the feature of regular items (Experiment 2).

As explained in the introduction, finding that category exceptions affect category decisions is surprising having in mind the dominant view in the literature that exceptions are represented separately from regular items. It is worth noting that this view is shaped by assumptions of models of category learning designed to predict learning and memory of one specific, different type of exceptions – exceptions that are highly confusable with the contrasting category members (Nosofsky, Palmeri, & McKinley, 1994; Love, Medin & Gureckis, 2004). Therefore, while this assumption may be useful in simulating the exceptions confusable with contrasting category items, there is no evidence that it generalizes more broadly to other types of exceptions. In contrast, our findings suggest that features of exceptions may be represented together with the features of regular items, and they may jointly affect generalization.

In addition to finding a robust evidence that category exceptions *can* expand category boundaries, we also report an interesting pattern that demonstrates the limits of this effect. When participants were asked to classify items that had probabilistic feature of Regular items (i.e. shape) and the rule of the Exception (i.e. color), they made category decision in accordance with the feature of the Exception. However, items that had probabilistic feature of Exception item and the rule of Regulars were classified based on the feature of the Regulars, both before and after Exception was introduced. Therefore, not any feature of Exception can expand category boundaries. This effect is specific to the dimension that is the most predictive of category membership based on the overall category structure.

Finding that exceptions stretch category boundaries only on the rule dimension is important. It suggests that not only that exceptions are not stored separately, but their representation is affected by the representation of regular items. Note that for exceptions in our experiments both features were equally, fully predictive. Therefore, the only reason to weigh these features differently could lie in the effect the category structure of regular items had on exceptions. This finding is in accordance with previously reported spill-over effect in memory representation of exception items (Savic & Sloutsky, 2019). Specifically, it was found that participants who tend to form rule-based representations of regular items also tend to form rule-based representations of exceptions of the same category. On the other hand, participants who form similarity-based representations of regulars, tend to form similarity-based representations of exceptions. Both of these findings run counter the view of regular and exception items being represented separately.

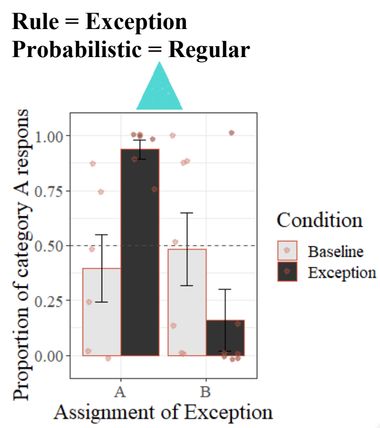

Figure 5: Generalization performance of participants with good memory for probabilistic features, collapsed across Experiments 1-2.

It is worth discussing potential alternative explanations of the current pattern of results. For example, one could wonder whether the reported pattern of generalization could be simply explained by participants learning only one dimension - the color of the stimuli. If participants did not learn the shape, then test items did not raise any conflict - participants simply always responded in accordance with the only dimension they've learned. Although this explanation would not speak against our main interpretation that features of exceptions are used in generalization, it would require modification of further interpretation – that the rule feature of the exception is preferable cue in categorization over the probabilistic feature of the regular items. Therefore, this is a potentially important concern. Therefore, we run a follow-up analyses on a subsample of participants from Experiment 1 and 2 and used as inclusion criteria memory data we did not report in our main analyses. To be selected participants had to have good memory (i.e. above chance level of .60) for shape of all of the training items. As shown in Figure 5, the pattern of generalization in this subsample is the same as the reported pattern for the whole sample. Therefore, it seems justified to conclude that participants did choose to rely on the rule of the exception over the probabilistic of the regular.

Conclusion

Across two experiments we found evidence that category exceptions change category boundaries and affect generalization of novel items. In addition, our findings suggest that the overall category structure also affects the representation of the exception items. Therefore, the current work suggests that when encountering information that violates one's expectations, adult human learners form a representation that is *flexible* enough to account for new evidence, but at the same time tends to be *stable* and preserves the previously learned structure of the category.

Building on this initial strong evidence, further research is needed in order to demonstrate whether the pattern we have found for learning simplified, artificial category structures holds for more complex category structures which may include other types of category exceptions.

Acknowledgments

This work was supported by the National Institutes of Health Grants R01HD078545 and P01HD080679 to Vladimir Sloutsky.

References

Castro, L., Savic, O., Navarro, V., Sloutsky, V. M., & Wasserman, E. A. (2020). Selective and distributed attention in human and pigeon category learning. *Cognition*, 204, 104350.

Deng, W.S., & Sloutsky, V.M. (2016) Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, 91, 24-62.

Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, 9(1), 160-168.

Gelman, S. A., & Markman, E. M. (1986). Categories and Induction in Young Children. *Cognition*, 23, 183-209.

Heit, E. (2000). Properties of Inductive Reasoning. *Psychonomic Bulletin & Review*, 7, 569-592.

Gelman, S. A., & Meyer, M. (2011). Child Categorization. Wiley Interdisciplinary Reviews: *Cognitive Science*, 2, 95-105.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning?. *Cognitive psychology*, 45(4), 482-553.

Lacroix, G. L., Giguere, G., & Larochelle, S. (2005). The origin of exemplar effects in rule-driven categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 272.

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

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

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.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition*, 34(1), 3-16.

Rips, L. J. (2001). Necessity and natural categories. *Psychological Bulletin*, 127, 827-852.

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

Savic, O., & Sloutsky, V. M. (2019). Assimilation of exceptions? Examining representations of regular and exceptional category members across development. Journal of *Experimental Psychology: General*, 148(6), 1071-1090.

Sloutsky, V. M. (2010). From Perceptual Categories to Concepts: What Develops? *Cognitive Science*, 34, 1244-1286.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.