Title
You Only Had to Ask Me Once: Long-term Retention Requires Direct Queries During Learning

Permalink
https://escholarship.org/uc/item/4x2945p5

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN
1069-7977

Authors
Love, Bradley
Sakamoto, Yasuaki

Publication Date
2009

Peer reviewed

You Only Had to Ask Me Once: Long-term Retention Requires Direct Queries During Learning

Yasuaki Sakamoto (ysakamot@stevens.edu)
School of Technology Management, Stevens Institute of Technology
Hoboken, NJ 07030 USA

Bradley C. Love (brad_love@mail.utexas.edu)
Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA

Abstract

What is the key to retaining learned study materials? We examine the role that direct queries play in the retention of category information acquired through inference learning. In inference learning, learners predict an unknown property of an item given its other properties and category membership. We manipulated query frequency across properties during inference learning and found that properties that were queried less often were remembered as well as properties that were queried more frequently. These effects extended from minutes to multiple day delays. Thus, asking about properties during inference training only a few times can immediately promote learning and greatly enhance the long-term retention of category knowledge. We situate our results in the broader memory and education literatures and consider how these findings constrain the development of category learning models.

Keywords: Retention; category learning; inference; direct query; explicit evaluation.

Introduction

What is the key to long-term retention of studied materials? In the present work, we investigate the role that direct queries play in category learning and recognition memory. The present work is unique in measuring retention over multiple day delays in category learning tasks. Retention has obvious importance in education and other everyday activities. However, the importance of consolidation to memory performance has been underappreciated by cognitive psychologists in general (Wixted, 2005) and grossly neglected by category learning researchers in particular. For instance, category learning studies that examine retention often impose delays of only a few minutes (e.g., Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004).

In addition to examining the role that direct query plays in consolidation, the current work advances our understanding of how category learning works using inference as the induction method. Typically, the classification induction method is used in category learning tasks. In classification learning, the learner is shown a stimulus and classifies it as belonging to one of a number of categories and then is supplied with corrective feedback. Like classification learning, inference learning is also a supervised learning method, but differs in that the category label is known and instead a missing feature value is inferred, followed by corrective feedback.

Figure 1: An inference learning trial in the current experiment is shown, in which the learner predicts an unknown property. The stimulus on the left differs from the stimulus on the right only on the values of the queried dimension (in this case, smaller vs. larger body size).

In category learning experiments using the inference method, the learner's goal is to correctly predict an unknown property of an item given the item's remaining properties and its category membership. Figure 1 displays an inference learning trial from the present experiment, in which the learner predicts whether the Tiger shark is small or large by guessing whether the left or right side describes the shark correctly. The left side of Figure 1 conveys the same information as the right side except for the queried dimension, size – the learner predicts the unknown property. The property that is queried in inference learning varies across trials. On the next inference learning trial, whether the Tiger shark has darker or lighter shade may be queried given the other properties and the category membership.

Inference learning can lead to more efficient learning of the categories and greater knowledge about the categories than classification learning (Sakamoto & Love, 2006; Yamauchi & Markman, 1998). In classification learning, the learner's goal is to successfully predict the category memberships of stimulus items. Unlike in inference learning, the category label is always queried in classification. Information about other perceptual properties is not queried in the classification procedure. Thus, classification learners try to discover which properties are useful for discriminating the members of different categories without being directly asked about these properties. In contrast, inference learners are directly asked about the properties associated with each category.

Sakamoto and Love (2006) have shown that classification learners display little knowledge about properties other than the property that is most useful in distinguishing members from different categories. On the other hand, inference learners acquired information about multiple properties that are queried during training, and this knowledge was retained after multiple day delays. Thus, querying played a critical role in the retention of knowledge by inference learners. However, it is unclear whether explicit querying is beneficial because it guides people's attention to the queried properties or because answering questions consolidates memory. These questions are addressed in the current experiment.

Although many studies have now examined inference learning (e.g., Anderson, Ross, & Chin-Parker, 2002; Chin-Parker & Ross, 2004; Colner, Rehder, & Hoffman, 2008; Johansen & Kruschke, 2005; Markman and Ross, 2002; Yamauchi & Markman, 1998), the processes underlying inference learning are still not well understood, at least not as well as those underlying classification learning. One open theoretical issue is the nature of retention in inference learning, in particular how consolidation is affected by explicit querying during training.

In the current work, we investigate the role direct query plays in the retention of category information under inference learning. We manipulate the frequency of query during inference learning. Does the retention of information about a dimension improve when the dimension is queried more often? Or is querying over certain frequency enough to preserve memory? One possibility is that repeatedly answering the queries serves as rehearsal and leads to improved memory. Then, more queries may result in better retention. Alternatively, direct query may improve retention by guiding people's attention to the property even on trials in which the property is not queried, as indicated in eye tracking studies of inference learning (Colner, Rehder, Hoffman, 2008). Once learners attend to the queried dimension, they develop an expectation that they will be asked again and attend to this dimension even when they are queried about another dimension. If this is the case, querying a few times may be enough to attract people's attention and improve their retention. We examine participants' memory about categories after a few minute and multiple day delays. We also address how error rate during learning affects retention.

In the Discussion section, we situate the present results, as well as inference and classification learning methods, within the broader memory and education literature concerned with the learning and retention of information. Inference and classification learning will be compared to direct and discovery problem solving methods (Klahr & Nigam, 2004). Furthermore, the role of queries in retention will be discussed in light of related test-based enhancement phenomena (see Roediger & Karpicke, 2006 for a review). Finally, we consider how the present findings constrain the development of category learning models.

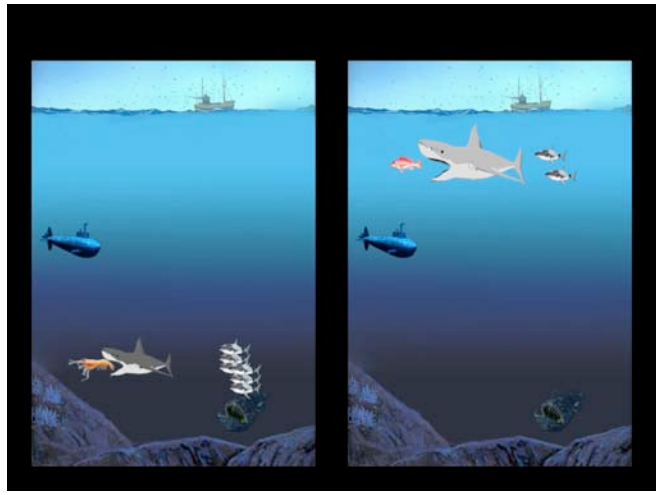

Figure 2: Snapshots of two sharks are shown side by side. The shark on the left, the prototype of the Sixgill shark, displays the opposite values on the five dimensions from the shark on the right, the prototype of the Tiger shark.

Method

Participants

Fifty University of Texas undergraduates completed the experiment.

Apparatus and Materials

Each undergraduate was run on an iMac with 17-inch display. The resolution was set 800 by 600 pixels.

The stimuli in the present experiment were animations of sharks swimming in the ocean. One animation cycle consisted of the shark appearing on the right side of the display, swimming to the left side, and disappearing when it reached the left edge. Each stimulus varied along the following five binary-valued dimensions: habitat (near the surface or bottom), diet (fish or shrimp), litter size (a few or many pups), body size (small or large), and shade (light or dark). Figure 2 displays snapshots of two sharks side by side with an opposite value on each dimension. The five dimensions were mapped randomly onto the logical structure shown in Table 1. For example, the first dimension was the habitat dimension for some participants, but it was the diet dimension for others. The dimension values were assigned according to the properties of the sharks used in the experiment.

Table 1: The abstract category structures used in the current experiment. Stimulus item A1, for example, belonged to Category A and had 5 perceptual dimensions. Category A might be Sixgill sharks. The first dimension might be size with value 1 indicating small and value 2 indicating large. Participants learned about the categories through inference. They predicted the values of the dimensions, one at a time, given the values of the other dimensions and the category label, with the constraint that the amodal category values (e.g., 2 in 11112) were never inferred.

Stimulus item	Dimension value
A1	11112
A2	11121
A3	11211
A4	12111
A5	21111
B1	22221
B2	22212
B3	22122
B4	21222
B5	12222

The categories contrasted Sixgill and Tiger sharks. Relative to the Tiger sharks, the Sixgill sharks are common in deep water (vs. surface to 340 m), often feed on shrimp (vs. fish), give birth to many pups (vs. from 10 to 80), are small (vs. 3 to 6 m), and have dark body shade (vs. grayish above and white below). Information about the sharks was collected from http://www.enchantedlearning.com and http://www.flmnh.ufl.edu/fish/Education/education.htm.

The categories in Table 1, like many natural categories, follow family-resemblance structures (Rosch & Mervis, 1975), in which properties tend to co-occur, but no single property is common to all members of a category. In Table 1, value 1 on each perceptual dimension signifies the value common to the Sixgill sharks when category A is the Sixgill shark. In this case, although most Sixgill sharks have value 1 on each dimension, occasionally they have value 2, which is common to the Tiger sharks. Participants were informed that the sharks vary in their properties and thus the two categories' members could display overlapping properties.

Design and Procedure

Undergraduates learned about the Sixgill and Tiger shark categories through inference. Each undergraduate completed two sessions. The initial session consisted of familiarization, training, filler, and test phases (described below). The participants received course credit for their participation in the initial session. Twelve to 33 days after completing the initial session (M = 23 days, Se = 1 day, median = 21 days), the participants completed a second session consisting of a retention phase, which was identical (except for the random presentation order) to the test phase they completed in the initial session. The participants received $7 for participating in the second session. The relatively wide range of delay is due to scheduling issues.

Familiarization Prior to learning about the sharks, participants were familiarized with the five stimulus dimensions. On each familiarization trial, a pair of sharks that differed on one of the five dimensions was presented, and the participants were asked to discriminate between the two possible values (e.g., "Which shark is larger? Left (Q) or right (P)?"). The participants pressed the P or Q key to indicate the right or left side is correct, respectively. After the participants responded, a blank screen was displayed for 1000 ms. Then they received visual (e.g., "Right! The correct answer is P." or "Wrong! The correct answer is Q.") and auditory corrective feedback (i.e., a low-pitch tone for errors and a high-pitch tone for correct responses), together with the correct shark (i.e., the foil disappeared) for one animation cycle. Then, a blank screen was displayed for 1000 ms and the next trial began. Participants completed 15 (3 per dimension) familiarization trials in a random order.

Training Following familiarization, participants completed a training phase, in which they learned about the shark categories in Table 1 through inference. On each training trial, the participants were shown two animated shark stimuli side by side as displayed in Figure 1. Whereas one stimulus correctly described the shark, the other stimulus did not. The correct and foil stimuli were randomly assigned to the left or right position. The two stimuli were identical to each other except for the value of a queried dimension; participants predicted a missing dimension value given the values of the other four dimensions and the category label. In Figure 1, for instance, the participant predicted the size of the presented Tiger shark.

We manipulated the frequency with which each of the five dimensions was queried. The first dimension was queried 24 times for the prototypical value, the second dimension 18 times, the third dimension 12 times, the fourth dimension 6 times, and the fifth dimension 0 time during training. The 60 training trials were broken down into three training blocks. The frequency of query for each dimension was distributed equally across the three training blocks.

Following the inference procedure used in most previous work (e.g., Chin-Parker & Ross, 2004; Yamauchi & Markman, 1998), the correct value of the queried dimension was always typical of the shark's category (e.g., value 1 for Sixgill and value 2 for Tiger). For example, the last perceptual dimension of item A1 (see Table 1) was never queried in the inference training trials because the correct value is inconsistent with the category-typical value.

The procedure in the training phase was identical to that in the familiarization phase except that the visual corrective feedback in the training phase specified the category membership (e.g., "Right! This is the Tiger shark.").

Filler Following training, the participants were shown a movie of 12 sharks swimming sequentially to prevent rehearsal of information from the training phase. Pictures of the Black-tip, Galapagos, Hammer Head, Horn, Lemon, Sandbar, Sharp Nose, Short-fin Maco, Whale, White, White-tip, and Zebra sharks were presented in a random order. Each shark was animated for 9000 ms with its name displayed at the bottom of the display, and a blank screen was displayed for 1000 ms before the next shark appeared.

Test Following the filler phase, the participants completed a test phase, in which their knowledge about the properties of the two shark categories from training was measured. The test phase consisted of a sequential presentation of 20 text questions, once in a random order, resulting in 20 trials. Ten forced-choice questions tested each of the 5 dimensions for the two categories. For example, the text "Tiger sharks:" was presented above the two choices "A: tend to be smaller" and "B: tend to be larger" when the size dimension of the Tiger shark was questioned. Another set of 10 questions was created in the opposite fashion. For instance, the text "tend to be larger:" was displayed above the choices "A: Tiger sharks" and "B: Sixgill sharks" when the shark associated with the larger size was questioned. The correct (i.e., category-typical) and foil choices were randomly assigned to the top (A) or bottom (B) position on each trial. Participants pressed the A or B key to indicate choice A or B is correct, respectively. No corrective feedback was given to participants to prevent learning during the test phase. After participants responded, a high-pitch tone sounded briefly and the text "Thank you" appeared beneath the choices for 2000 ms. Then, a blank screen was displayed for 2000 ms and the next trial began.

Retention In the second session, the participants completed a retention phase. The procedure for the retention phase was identical to that for the test phase from the initial session. The retention phase measured how well participants retained information about the shark categories learned in the initial session after a multi-day delay (i.e., 12 to 33 days).

Results

All participants were included in the analyses. Our main interests are the participants' performances on each dimension in the training, test, and retention phases. Figure 3 summarizes the results from the present experiment.

Training

Participants' training accuracies for dimensions queried 6, 12, 18, and 24 times did not differ significantly, $F(3, 147) = 2.11$, $MSe = .03$, $p = .1$. The lack of significant differences in training accuracies between dimensions queried 6 times and 24 times ($t(49) = -1.63$, $p = .11$) suggest that learning takes place quickly in inference.

Test

Participants' test accuracies for dimensions queried 0, 6, 12, 18, and 24 times differed significantly, $F(4, 196) = 13.24$, $MSe = .09$, $p < .01$, partial $\eta^2 = .21$. When only the dimensions that were queried were compared (6, 12, 18, and 24), the differences in the participants' test accuracies did not reach significance, $F(3, 147) = 2.36$, $MSe = .07$, $p = .07$.

As can be seen in Figure 3, the participants did not learn about the non-queried dimension. Whereas their test accuracy on the dimension that was queried 0 time did not differ significantly from the chance level of .5 ($p = .8$), their test accuracy on each of the other dimensions that were queried did ($p < .01$ for each).

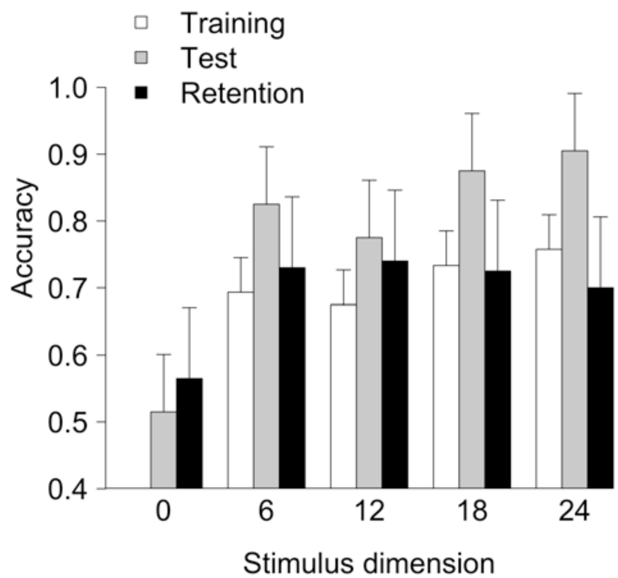

Figure 3: Undergraduates' performances are shown in the training, test, and retention phase of the experiment. Stimulus dimension indicates the number of total queries during training. There are no training data for the dimension that was not queried. Error bars represent the upper bounds of the 95% confidence intervals (Loftus & Masson, 1994).

Retention

Participants' retention accuracies for dimensions queried 0, 6, 12, 18, and 24 times did not differ significantly, $F(4, 196) = 1.89$, $MSe = .14$, $p = .11$. Mirroring test phase performance, whereas performance on queried dimensions was above chance ($p < .01$ for each), performance on the non-queried dimension did not differ significantly from chance ($p = .29$). Whether a dimension was queried many times or only a few times did not have a strong effect on inference learners' test and retention performances. As shown in Figure 3, for queried dimensions, the participants' accuracies during test (78%) and retention (69%) appear to be unaffected by the number of queries during training.

Relationship between Training and Later Performances

Training accuracy was positively correlated with the test accuracy for each queried dimension ($r = .39$, $p < .01$ for the dimension queried 6 times; $r = .43$, $p < .01$ for 12; $r = .41$, $p < .01$ for 18; $r = .49$, $p < .01$ for 24). Training accuracy was also positively correlated with the retention accuracy on each queried dimension except for the one queried least frequently (r = .18, p = .22 for the dimension queried 6 times; r = .46, p < .01 for 12; r = .3, p < .04 for 18; r = .35, p = .01 for 24). Making errors thus correlates with worse test and retention performances.

Discussion

The current study examined how explicit queries during inference training shape category acquisition and retention. We manipulated how often the dimensions were queried during inference training using categories with a family-resemblance structure. Explicit query resulted in improved retention performances regardless of the frequency of query. Whereas querying 6 times resulted in above chance retention performance that was statistically equal to querying 24 times, not querying at all resulted in a chance performance. Participants appeared to master the values of queried dimensions after a few training trials, suggesting that only a few queries are necessary to acquire and retain category-property relationships and that there is little or no benefit of additional queries.

The lack of a performance difference between infrequently and frequently queried dimensions suggests that any query of a dimension serves as a signal to the learner to attend to and encode category-property relationships for that dimension, even on trials in which the dimension is not queried. Once the learners are queried about a dimension, they develop an expectation that they will be asked again and attend to the information about the dimension even when they are asked about another dimension. Such a mechanism is consistent with the eye-tracking studies showing that inference learners look at dimensions that are not directly asked during training (cf. Colner et al., 2008), and can account for the high test and retention performances for the queried dimensions but the low performances for the non-queried dimension in the current experiment.

Another interesting finding from the current experiment was that making errors was correlated with worse test and retention performances. This result seems to contradict the finding that conditions that are more difficult at study often lead to better performance at test (e.g., Bjork, 1994). One possibility is that inference learning does not result in many errors that can lead to source monitoring problems, in which the learners confuse response with feedback, such as "I responded large but it was small, or was it the other way around?". Another possibility is that some stimuli are simply more memorable for certain participants and this effect drives performance across phases.

Whereas the inference learners in the present experiment were at chance level in test and retention for dimensions that were not queried during training, pervious work examining inference learning has found that inference learners acquire information about both queried and non-queried properties (Anderson et al., 2002; Sakamoto & Love, 2006). Inference learners were not focusing exclusively on the queried dimensions during training. However, after multiple day delay, inference learners did not retain the information about non-queried dimensions (Sakamoto & Love, 2006), consistent with the current finding. One explanation for the lack of learning about the non-queried dimension in the current work is task difficulty. Capacity limitations may have prevented learners from entertaining the non-queried dimension in the present work. Whereas there was only one non-queried dimension and four others queried in the present experiment, there were two non-queried dimensions and two (Anderson et al., 2002) or three queried (Sakamoto & Love, 2006) in the previous work. Further, remembering information about the dimensions that were queried less frequently in the current work might have consumed additional cognitive resource, thus preventing the participants from attending to the non-queried dimension. Indeed, the inference learners' training and test performances suggest that the inference task in the present experiment was more demanding than that in the previous experiment (.72 for training and .78 for test in the present experiment vs. .94 for training and .88 for test in Sakamoto & Love, 2006).

Analogous to the inference result in which only queried dimensions, which are explicitly evaluated, are retained, classification learners tend to only retain information about the dimension that is diagnostic in discriminating members of different categories (Sakamoto & Love, 2006; but see Bott, Hoffman, & Murphy, 2007). Classification learners actively engage in hypothesis testing involving the diagnostic dimension when they predict the category labels (e.g., Nosofsky, Palmeri, & McKinley, 1994; Sakamoto & Love, 2004), and this explicit evaluation consolidates memory. Thus, people retain information that they explicitly evaluate, and direct queries can facilitate this process.

Implications for Education and Models of Category Learning

From the standpoint of learning the category-property association, inference is similar to direct instruction, and classification is more like discovery learning. Whereas inference learners are explicitly asked about properties associated with the category, classification learners are only asked about the category membership and have to discover the properties on their own. More efficient learning of the categories and greater knowledge about the categories in inference learning than classification learning may be related to the findings that direct instruction can lead to successful learning by many more children than discovery learning (Klahr & Nigam, 2004). Although discovery learning may require more active processing, it can lead to errors and confusion as many classification learners experience. In contrast, direct instruction, like the inference procedure, can make the task easier and guide the learner on what needs to be acquired. The lack of guidance may make classification less efficient like discovery learning.

Direct query can be beneficial to promoting learning and establishing memory. Moreover, querying a few times can be as effective as querying numerous times. Asking about properties during inference training, even only a few times, can immediately promote learning and greatly enhance the learner's retention of knowledge of category properties. These results parallel the findings that testing on material can be more beneficial to establishing memories than additional study (e.g., Roediger & Karpicke, 2006). The present results suggest that frequency of testing may not matter even for long-term retention of learned materials.

Existing category learning models do not specify how queries shape retention and how knowledge is consolidated (cf. Sakamoto & Matsuka, 2007). Thus our results provide guidance for the further development of these models. For example, models need to address the role queries play in shaping attention and consider that a property need only be queried a few times to manifest the full benefits of inference learning and to result in long-term retention.

Many theories of category learning assume that errors play a central role in learning (e.g., Kruschke, 1992; Love et al., 2004). These theories are guided by conditioning phenomena, which suggest that errors are necessary for changes in memory (e.g., Rescorla & Wagner, 1972). Errors mediate memory storage by leading to greater focus on error-producing items (e.g., Mackintosh, 1975). These category-learning models need to be elaborated to address our finding that more errors can be associated with worse learning and retention.

Acknowledgments

This work was supported by AFOSR FA9550-07-1-0178 and NSF CAREER #0349101 to B.C.L.. We thank Tyler Davis for his help in collecting data.

References

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

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General*, *136*, 685–699.

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

Colner, B., Rehder, B., & Hoffman, A. B. (2008). Feature inference and eyetracking. In V. Sloutsky, B. Love, K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1170–1175). Mahwah, NJ: Erlbaum.

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

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*, 661–667.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

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

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimulus with reinforcement. *Psychological Review*, *82*, 276–298.

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

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.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

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). Sizable sharks swim swiftly: Learning correlations through inference in a classroom setting. In R. Sun and N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Vancouver, Canada: Cognitive Science Society.

Sakamoto, Y., & Matsuka, T. (2007). Incorporating forgetting in a category learning model. In *Proceedings of International Joint Conference on Neural Networks*.

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science*, *14*, 6–9 .

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