Variability in category learning: The Effect of Context Change and Item Variation on Knowledge Generalization

Dustin D. Finch (ddfinch@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th St.
Bloomington, IN, 47401 USA

Paulo F. Carvalho (pcarvalh@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th St.
Bloomington, IN, 47401 USA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th St.
Bloomington, IN, 47401 USA

Abstract

We explore how context change and item variation during natural category learning influence memory and generalization to new examples. Participants studied either images of the same bird or varied birds from each of several categories. These images could be presented in a constant background color or different background colors. During test, birds were presented in only one of the studied background colors. Performance at test depended on the context overlap between study and test, with better performance when there was minimal context change during study. Also, contrary to previous findings, we found that learners generalized better when items were repeated during study and remembered old items better when items were varied during study. When there is a moderate degree of context change, there is no benefit of repetition or variation for either novel or old items. These results indicate that context change and item variation have complementary effects on learning.

Keywords: concept learning, memory, context, variability, repetition.

Introduction

In our every-day life, we have all recognized someone without being able to "place" the person or remember their name. For example, even though we may have been visiting their office for years, we might fail to remember the name of our dentist when we see her at the grocery store. A related feeling is reported by students in preparation for exams. They are able to effectively remember the content the night before the exam in their dorm room, but that knowledge seems to have vanished when they are tested in an exam context different from their dorm room study context.

These are real-life examples of a property of our memory system – context dependency. When trying to remember something, reinstating the conditions of the encoding situation (i.e., the context) helps retrieve the memory. For example, Godden and Baddeley (1975; see also, Godden & Baddeley, 1980; Smith, Glenberg, & Bjork, 1978) had people learn a list of words in one of two contexts: either underwater or on land. When asked to recall this list, people recalled more words when the testing context was the same as the study context (e.g., study done underwater and recall done underwater as well), compared to when the testing context was different (e.g., study done underwater and recall done on land). This effect has been repeatedly found across different tasks, paradigms and labs (see Smith, 2013 for a recent review and Smith & Vela, 2001 for a meta-analyses), including educational settings (Abernethy, 1940; but see Saufley, Otaka, & Bavaresco, 1985).

Why is memory context dependent? One leading theory is that as memories are created, background information (e.g., how you feel, the smell, sounds, or the room you are in) is encoded into that memory. This background information can then be used as cues to retrieve that memory (Hintzman, 1986; Bjork and Bjork, 1992; Smith 2013). Reinstating the study conditions at test makes these cues available, improving recall.

However, context dependency can have negative consequences, because a specific memory might become too tied to a specific context. For example, a student might study the mitochondrial structure in their room while listening to Pearl Jam. When later testing themselves on the mitochondrial structure while listening to Pearl Jam and still in their room, they are able to successfully retrieve the information studied. However, when asked to write about the process in a quiet classroom surrounded by 30 other students, retrieving this memory might be difficult. One possible solution to this issue is to decontextualize the specific memory.

Memory is considered to be decontextualized when removing the contextual cues associated with it does not affect retrieval. As the example above shows, decontextualization is particularly important when the test situation is different from the study situation. One way to do this is by varying study context and then testing people's memory in a new context (Smith et al., 1978; Smith, 1979; Smith & Handy, 2014; for evidence in classroom settings see Smith & Rothkopf, 1984). Smith et al. (1978) tested this by having a group of people learn a list of words in two separate rooms and a different group who studied the words in one room only (both groups of learners were asked to leave the room and return during learning). All participants were tested in a novel room. People recalled more words when they studied in two different rooms compared to studying in one room only. Studying in two different contexts led to the creation of decontextualized memories that were more easily retrieved in a new context.

Thus, previous research has demonstrated that our memory includes information about the learning situation, which can have a positive or negative effect on later retrieval. A possible explanation for this effect is derived from the MINERVA 2 model (Hintzman, 1986). This model suggests that every memory is encoded into its own memory trace which includes both target and context information. When accessing previous memories, common properties between the traces will be more salient at retrieval. In this way, when information is studied in one context only, all encoded information (both contextual information as well as target content) are highly salient at retrieval. Conversely, if the target information is studied across several contexts, it will be more salient because it is the only common information across all contexts. This will contribute to good recall, even when context cues are not available at retrieval.

A similar idea has been proposed in the context of generalization of category learning to new items (Hintzman, 1984). A goal of most category learning is to extend a category learning experience to novel items (Goldstone, Kersten, Carvalho, 2013). One way to improve category generalization is by increasing the variation across the studied items. When people study a category by seeing several different examples of the category, they are better at generalizing this knowledge to new items of the category compared to when they see only a few items repeated (controlling for total exposure time; e.g., Homa, Cross, Cornell, Goldman, & Shwartz, 1973; Posner & Keele, 1968; Wahlheim, Finn, & Jacoby, 2012). For example, learning the category dog by seeing examples of *German Shepherds, Bulldogs, Chihuahuas,* and *Poodles* would result in better generalization to new members of the category dog (e.g., *Pugs*) compared to studying only examples of *Chihuahuas*. Seeing several different items allows people to notice and extract the overlapping properties among all examples. This allows for better subsequent generalization (Hintzman, 1984).

In sum, both context change and item variation can be seen as different, perhaps complementary, ways of increasing study variability with positive results. Here we propose to extend previous research on context dependent memory and item variation to a situation where generalization of the knowledge is required. We analyze two main questions: (1) can context change improve not only memory retrieval but also knowledge generalization to novel items, and (2) can context change and item variation work together to result in improved generalization or is this high level of combined variability detrimental for memory and generalization?

We analyze these two questions by having learners learn bird species by studying examples of each species and manipulating the number of different items studied as well as the study context of the items. According to previous research with memory tasks, and the predictions of MINERVA 2, we expect to see better generalization in situations for which there is a high congruency between study and test contexts. Moreover, previous research suggests that both item variation and context variation will promote better generalization to new items at test, by increasing study variability and the detection of the category-relevant properties shared by members of a category (Hintzman, 1984).

An Experiment

In this experiment participants learned twelve different species of birds by studying examples of each species. We manipulated the context of study by changing the color of the background on which the bird was presented. In addition to manipulating the context of study we also manipulated the amount of item variation presented within each category.

Methods

Participants. A total of 238 undergraduate students at Indiana University volunteered to participate in this experiment in return for partial course credit. Participants were randomly assigned to one of the following conditions: No Context Change ($N = 73$), Medium Context Change ($N = 82$), and High Context Change ($N = 83$).

Figure 1: Examples of the birds shown during the study. The first row shows the same bird with the same background color, illustrating what participants would see in the Repetition and No Context Change Condition. The second row shows a different bird with the same background color, illustrating the Variation and No Context Change Condition. The third row shows a different bird in two different background colors, illustrating the Variation and Medium Context Change Condition. The last row shows different birds with different background colors, illustrating the Variation and High Context Change condition.

Stimuli & Materials. The stimuli were 72 images of birds from twelve different species (6 images from each bird species; see Figure 1 for examples), available as part of a database of bird images (Wahlheim, Dunlosky, & Jacoby, 2011). We selected the species Black-headed Grosbeak (*P. melanocephalus*), Blue Jay (*C. cristata*), American Goldfinch (*S. tristis*), American Tree Sparrow (*S. arborea*), Bank Swallow (*R. riparia*), American Robin (*T. migratorius*), Blue Bunting (*C. parellina*), Clark's Nutcracker (*N. columbiana*), Brambling (*F. montifringilla*), Bachman's Sparrow (*P. aestivalis*), Brown-chested Martin (*P. tapera*), and Bluethroat (*L. svecica*).

During the task, participants classified the images using the scientific classification of the species. We created four different versions of each image by changing the background color to one of the following: red, green, blue, pink (see Figure 1). Thirty different images were chosen to be presented during the study phase, and the remaining 42 images were presented only during test. Which images were chosen to be used during study and test was counterbalanced across participants. The stimuli were presented and responses from the participates were collected using a web-browser and the experiment software Collector (a free and open-source platform to run experiments using a web-browser available from https://github.com/gikeymarcia/Collector).

Procedure. An overview of the procedure used in this Experiment is presented in Figure 2. There were two phases to the experiment: Study and Test. Participants completed 48 study trials and 72 test trials.

During the study phase a picture of a bird and its species name would appear on the screen for eight seconds. Afterwards, the name was replaced by a Judgment of Learning prompt. Participants were asked to provide a judgment of learning on a 0 (0% confidence of correct classification)-100 (100% confidence of correct classification) scale and were told to try to use the whole scale while providing their judgments.

Immediately following the Study phase participants played Tetris for one minute until starting the test phase. During the Test phase, on each trial a bird was presented in the center of the screen, along with the name of the twelve species participants had studied. Participants were asked to click on the species name to which they believed the bird belonged. No feedback was given during the test.

For study and test, one third of the participants saw the images presented to them always in the same background color (No Context Change). The color of the background was counterbalanced across participants. Another third of the participants saw the images in two background colors (Medium Context Change). Half of the items in each category were in one color (e.g., red) and the other half in another color (e.g., green). Which two colors were selected was counterbalanced across participants. The last third of the participants always had a new background color for the items studied in each category (High Context Change). Participants never saw a repeated background color within each category, but colors were repeated across categories. Test items were presented always in the same background color – a color chosen from one of the studied colors and counterbalanced across participants.

During study we manipulated the number of different items participants saw for each category. For six out of the twelve categories, one item from each category was repeated four times (Repetition Condition). The other six categories had item variation in which each category had four different studied items (Variation Condition). Which item was repeated was counterbalanced across participants.

During Test we manipulated whether the item was old (studied) or was novel. For categories studied in the Repetition Condition participants saw one old item and five novel items. For categories studied in the Variation Condition, participants saw four old items and two novel items. This was done in order to keep constant the total number of test items across conditions. For both the study and test phases, trials were presented in random order.

		Study Phase				Old Items Test				Novel Items Test				
Context Change Condition	Variation Condition	Pres. 1	Pres. 2	Pres. 3	Pres. 4	Item 1	Item 2	Item 3	Item 4	Item 1	Item 2	Item 3	Item 4	Item 5
No Context Change	Repetition	Ex.1	Ex.1	Ex.1	Ex.1	Ex.1				Ex.2	Ex.3	Ex.4	Ex.5	Ex.6
	Variation	Ex.1	Ex.2	Ex.3	Ex.4	Ex.1	Ex.2	Ex.3	Ex.4				Ex.5	Ex.6
Medium Context Change	Repetition	Ex.1	Ex.1	Ex.1	Ex.1	Ex.1				Ex.2	Ex.3	Ex.4	Ex.5	Ex.6
	Variation	Ex.1	Ex.2	Ex.3	Ex.4	Ex.1	Ex.2	Ex.3	Ex.4				Ex.5	Ex.6
High Context Change	Repetition	Ex.1	Ex.1	Ex.1	Ex.1	Ex.1				Ex.2	Ex.3	Ex.4	Ex.5	Ex.6
	Variation	Ex.1	Ex.2	Ex.3	Ex.4	Ex.1	Ex.2	Ex.3	Ex.4				Ex.5	Ex.6

Figure 2: Schematic representation of the experimental design for study of the categories in each of the study conditions. Each square represents one example of a bird in the species. The color of the square represents the background color in which the bird was presented.

Results

Judgments of Learning Participants were more confident on their ability to correctly classify repeated items (M = 53.48, SD = 22.46) at test compared to varied items (M = 47.92, SD = 21.48), F (1,235) = 38.61, p < .0001. Moreover, participants' Judgments of Learning for repeated and varied categories were influenced by the degree of context change presented, F (2,235) = 5.52, p = .005. When participants studied in the No Context Change and Medium Context Change conditions, they were more confident in their ability to correctly classify repeated items (M = 55.91, SD = 21.06, and M = 55.95, SD = 22.71, respectively; both ps < .0001) compared to varied items (M = 47.59, SD = 20.80, and M = 48.80, SD = 21.85). However, when there was High Context Change, participants rated their ability to later classify both types of items equally (M = 48.90, SD = 22.94, and M = 47.35, SD = 21.93 for repeated and varied items, respectively; p = .349).

Test Performance We started by looking at the effect that changing backgrounds had on overall test performance (see Figure 3). The results show that the amount of Context Change had an overall effect on test performance, F (2,235) = 3.61, p = .029. Planned comparison tests indicate that studying items with four different backgrounds (M = 65.27, SD = 30.31) worsens performance compared to two backgrounds (M = 69.41, SD = 30.62; p < .0001), or one background (M = 69.76, SD = 30.41; p = .001). Participants' performance was equivalent when comparing one background to two backgrounds (p = .531).

Moreover, participants are better at classifying old items (M = 0.51, SD = 0.50), compared to novel items (M = 0.43, SD = 0.50), F (1,235) = 97.09, p < .0001 (see Figure 3). Importantly, item variation and repetition had different effects for old and novel items, F (1,235) = 42.36, p < .0001. When classifying novel items, participants are better for categories studied with item repetition (M = 0.45, SD = 0.50) compared to categories studied with item variation (M = 0.41, SD = 0.49; p = .026). Conversely, the opposite pattern is seen for old items (M = 0.47, SD = 0.50, and M = 0.54, SD = 0.50, for repetition and varition respectively; p < .0001).

Lastly, the effect of item variation depends on the background change condition and whether the stimulus is old or novel, F (2,235) = 14.14, p < .0001. To investigate this three-way interaction, we calculated the difference between performance for categories studied with variation and categories studied with repetition (Variation – Repetition). We used this score to investigate differences when classifying old and novel items and among different background change conditions (see Figure 4).

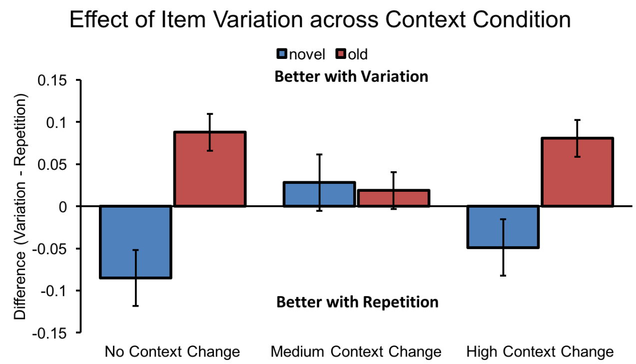

Figure 4: Difference in test performance for novel and old items between categories studied with variation and categories studied with repetition (Variation – Repetition). Positive values represent better categorization with item variation while negative values represent better categorization with item repetition. Error bars represent standard errors of the mean.

The results indicate that when studying items in the No Context Change or High Context Change conditions, pariticipants performed better at classifying novel items for categories studied with repetition, while for old items they performed better for categories studied with variation (both ps < .0001). However, when studying the species in the Medium Context Change condition, participants' performance was equivalent for old and novel items (p = .720), and equivalent for categories studied with repetition and variation for both old (t-test comparing with 0: t (81) = 0.72, p = .469), and novel items (t-test comparing with 0: t (81) = 1.08, p = .282).

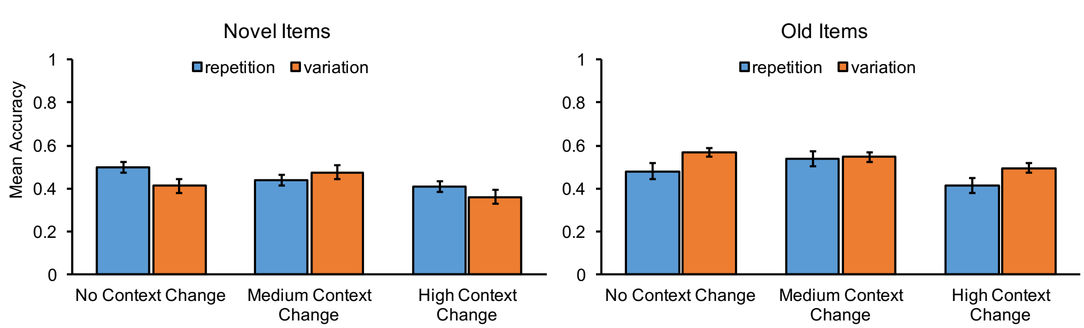

Figure 3: Performance during the test phase by amount of context change during study and item novely at test. Change performance in task is 0.08. Error bars represent standard errors of the mean.

Discussion

In this study we investigated how context change and item variation during study of natural categories (bird species) influence later memory and generalization to new examples.

We found evidence for context dependency in a generalization task. Learners performed better when there was No Context Change and Medium Context Change. As the overlap between study and test conditions (background colors) increased so did test categorization performance. For No Context Change and Medium Context Change conditions there was a high degree of overlap between study and test contexts. For High Context Change there is only a small degree of overlap between study and test. To the best of our knowledge, this is the first demonstration of this context reinstatement effect in a categorization task, and is in agreement with previous research showing a positive relationship between increased overlap of study and test conditions and memory performance (e.g., Smith et al., 2014). Moreover, it extends previous evidence to performance in classification of novel items (generalization) that go beyond remembering specific instances.

As mentioned in the introduction, previous research has shown that adding item variation improves both recognition memory for old items placed in new contexts and categorization of novel items. (e.g., Wahlheim et al., 2012). However, contrary to previous evidence, here we found that learners generalized better when items were repeated during study and remembered the categories of old items better when items were varied during study. There are several reasons for these results. It is possible that the context change manipulation and the fact that participants are studying categories with and without variation simultaneously might have impacted the results. It is possible that the change in contexts lead learners to try to group items into categories by background color, leading to overall lower attention to the similarities between items of the same category but with different background colors. However, the fact that we saw similar results for participants who studied the categories in one context only indicates that this might not be the only factor influencing these results.

Another possibility is related to the high degree of similarity within items of each category and between the studied items and the novel items presented at test. It has been proposed before that the benefits of item variation during study are related to the degree of dissimilarity within the category and between the studied and test items (Hahn, Bailey, & Elvin, 2005; Stewart & Chater, 2002). In our study, the items within each category were highly similar (see Figure 1). Moreover, the novel items used at test were very similar to the studied items as well. This high level of similarity might have reduced the possible benefits of increased item variation at study, because it is relatively easy to identify similarities across items of the same category during study and to extend this knowledge to very similar items at test. It is possible that if more dissimilar items had been used at test and/or during study the results would have been different. Post-hoc analyses looking at the effect of item similarity on test performance suggest that this might be the case. We used an indirect measure of category similarity, how confusable a category is with another category, based on participants' Reponses to compare similar and dissimilar categories. When we compared test performance between categories frequently confused with another category (e.g., *C. parallina* and *C. cristata*) and categories that were not confused with each other (e.g., *L. svecica and N. columbiana*), we see that the pattern of results described is reversed only when the categories are easier to discriminate (not as frequently confused). For these categories, participants remember better the categories of old items repeated during study and generalized the categories to novel items better when items were varied during study.

It is, however, interesting to note that item variation improved memory for old items in this experiment relative to repeating the old items. To our knowledge, this is the first demonstration of such an effect. Item variation might have worked to promote more efficient episodic memory for each of the items presented, making them more distinct and therefore more memorable (Nairne, 2006; Rawson & Van Overschelde, 2008; Schmidt, 1985; von Restorff, 1933). Congruent with this hypothesis, the benefit of item variation is numerically larger when there is No Context Change during study, compared to when there is High Context Change, perhaps because context change in itself would also work to make items more memorable by increasing attention to the item. However, learners seem to be unaware of the potential benefits of variation, generally showing higher degrees of confidence in their ability to later categorize items studied repeatedly than varied items, similar to what has been shown in previous research (e.g., Wahlheim et al., 2012).

Finally, the relative benefit of item *repetition* compared to item *variation* when classifying old and novel items at test is modulated by the amount of context change present during study. We found that when there is a medium degree of context change (i.e., half the trials were presented in one context and the other half in another context), there is no benefit of repetition or variation for either novel or old items. One possible reason for these results is that the partial context overlap led participants to ignore the context changes occurring as they presented a source of confusability, focusing instead on the properties of the items presented. In these conditions, both repetition and variation would work to promote better episodic memory allowing for an efficient generalization at test as well.

Overall, these results are preliminary evidence that context change and item variation influence learning with consequences for knowledge generalization. Moreover, it provides initial evidence of an interaction between the effects of context change and item variation during study on learners' ability to subsequently generalize their knowledge. An effect that students seem unware of. This has potential implications for educational settings, where context change (e.g., lecture classroom, laboratory section, library) and the examples given during study (e.g., different examples of the same concept or repeated examples across contexts) are frequent occurrences that can be combined for the best learning.

Acknowledgments

This research was supported in part by National Science Foundation REESE grant 0910218 and Department of Education IES grant R305A1100060 to RLG, and Graduate Fellowship SFRH/BD/78083/2011 from the Portuguese Foundation for Science and Technology (FCT) to PFC. Stimuli images courtesy of the Aging, Memory, & Cognitive Control Lab (http://psych.wustl.edu/amcclab/AMCC). We thank Alifya Saify for her help with stimuli creation.

References

Abernethy, E. M. (1940). The effect of changed environmental conditions upon the results of college examinations. *The Journal of Psychology, 10*(2), 293-301.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From learning processes to cognitive processes: Essays in honor of William K. Estes, 2*, 35-67.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Shwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101*(1), 116.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology, 66*(3), 325-331.

Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British Journal of psychology, 71*(1), 99-104.

Goldstone, R. L., Kersten, A., Carvalho, P. F. (2013). Concepts and Categorization. In Healy, A. & Proctor, R. *Handbook of Psychology Vol. 4: Experimental Psychology* (pp. 607-630). New York City, NY: Wiley.

Hahn, U., Bailey, T. M., & Elvin, L. B. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition, 33*(2), 289-302.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(2), 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. *Distinctiveness and memory*, 27-46.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77(3p1)*, 353.

Rawson, K. A., & Van Overschelde, J. P. (2008). How does knowledge promote memory? The distinctiveness theory of skilled memory. *Journal of Memory and Language, 58*(3), 646-668.

Saufley, W. H., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: Classroom tests and context independence. *Memory & Cognition, 13*(6), 522-528.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 565-578.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5*(5), 460.

Smith, S. M. (2013). Effects of environmental context on human memory. In T. Perfect & D. Lindsay (Eds.). *The SAGE Handbook of Applied Memory* (pp. 162-182). London, UK: Sage Publications.

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1582.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*(3), 341-358.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review, 8*(2), 203-220.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*(4), 342-353.

Smith, S. M., Handy, J. D., Angello, G., & Manzano, I. (2014). Effects of similarity on environmental context cueing. *Memory, 22*(5), 493-508.

Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 893.

Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld [Analysis of processes in the memory trace: 1. On the effect of group formations on the memory trace]. *Psychologische Forschung, 18*(1), 299-342.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750-763.

Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: evidence for variability neglect. *Memory & Cognition, 40*(5), 703-716.