Study preferences for exemplar variability in self-regulated category learning

By: Christopher N. Wahlheim, K. Andrew DeSoto

Wahlheim, C. N., & DeSoto, K. A. (2017). Study preferences for exemplar variability in self-regulated category learning. *Memory, 25(2)*, 231-243. doi: 10.1080/09658211.2016.1152378

This is an Accepted Manuscript of an article published by Taylor & Francis in *Memory* on 25 Feb 2016, available online: http://www.tandfonline.com/10.1080/09658211.2016.1152378

Abstract:

Increasing exemplar variability during category learning can enhance classification of novel exemplars from studied categories. Four experiments examined whether participants preferred variability when making study choices with the goal of later classifying novel exemplars. In Experiments 1–3, participants were familiarised with exemplars of birds from multiple categories prior to making category-level assessments of learning and subsequent choices about whether to receive more variability or repetitions of exemplars during study. After study, participants classified novel exemplars from studied categories. The majority of participants showed a consistent preference for variability in their study, but choices were not related to category-level assessments of learning. Experiment 4 provided evidence that study preferences were based primarily on theoretical beliefs in that most participants indicated a preference for variability on questionnaires that did not include prior experience with exemplars. Potential directions for theoretical development and applications to education are discussed.

Keywords: Category learning | learning strategies | metacognition | self-regulated learning

Article:

Category learning is a fundamental aspect of cognition that simplifies the environment and allows for flexible adaptation of known information to novel circumstances. Effective regulation of category learning is paramount for many everyday tasks. For example, a participant attempting to classify novel species of birds would benefit from choosing to study a variety of examples. Despite the central role of category learning in everyday activities, little is known about how individuals regulate such learning. The present investigation was conducted to examine the extent to which participants choose to incorporate variability into their study of category exemplars, the metacognitive bases for those choices, and the outcomes of those choices.

Studies of self-regulated learning typically have examined strategies that people use to learn verbal materials (for reviews, see Bjork, Dunlosky, & Kornell, 2013; Kornell & Finn, in press). Research on the allocation of study time has shown that participants study information longer when perceived learning is low (Dunlosky & Hertzog, 1998), but only for information they expect to remember (Metcalfe & Kornell, 2003, 2005). In addition, research on study choices has shown that participants prefer strategies that enhance learning, such as self-testing with feedback (e.g., Kornell & Son, 2009) and distributed study (e.g., Son & Kornell, 2009). However, participants sometimes do not take full advantage of testing opportunities (e.g., Karpicke, 2009) nor do they always comprehend that distributed study is more effective at longer lags (e.g., Wissman, Rawson, & Pyc, 2012). Thus, although participants may be able to enhance their performance by choosing appropriate strategies, they are sometimes unaware of the most effective way to utilise those strategies.

Despite the many investigations of self-regulated learning, only one recent study has examined study choices in category learning. Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) examined whether participants preferred blocked or intermixed study of natural category exemplars (i.e., bird species). Studies have shown that novel exemplars are classified more accurately following intermixed than blocked study (for a review, see Rohrer, 2012), but participants are not always aware of this advantage. For example, most participants reported believing that novel classification was better following blocked than intermixed study after classifying paintings from different artists, despite the fact that intermixed study produced better performance (Kornell & Bjork, 2008). Moreover, Tauber et al. (2013) discovered that most participants preferred to block their study of birds from various families. This preference occurred even though participants in a separate study that included the same materials predicted better performance on novel exemplar classification after intermixed rather than blocked study (Wahlheim, Dunlosky, & Jacoby, 2011). Together, these findings are consistent with the notion that participants do not always comprehend the effectiveness of certain study strategies.

Similar to intermixed study, increasing the variability of studied exemplars by including a greater number of distinct instances can also enhance classification of novel exemplars (cf. Posner & Keele, 1968). For example, Wahlheim, Finn, and Jacoby (2012) showed that studying six exemplars twice enhanced classification of novel bird species relative to studying two exemplars six times. When participants were administered a questionnaire either before or after the classification test, they estimated that variability was a more effective study strategy than repetitions. However, participants did not estimate that variability would be better than repetitions (also see Doyle & Hourihan, 2015) when making judgements for individual categories following study and just prior to the test (called category learning judgements, or CLJs). Given these inconsistencies, the extent to which participants prefer variability when studying natural categories for a test of novel exemplars is an open issue.

The aim of the present study was to address this issue by examining participants' preferences for variability when studying exemplars of bird families for a classification test of novel exemplars from the same families. A primary determinant of study preferences is the metacognitive bases on which choices are made. A direct relationship between metacognitive judgements and study choices has been shown in the learning of verbal materials (e.g., Koriat, Ma'ayan, & Nussinson, 2006), and this relationship might also be present in category learning situations. However, study choices can be based on multiple forms of metacognitive judgements, with two primary bases for choices being recent experiences and overarching theoretical beliefs (e.g., Kelley & Jacoby, 1996; Koriat, 1997). Experience-based choices are influenced by recent or current experiences, such as the fluency that is produced by the size of stimuli (e.g., Rhodes & Castel, 2008), whereas theory-based choices access declarative knowledge about the effectiveness of study strategies (also see Dunlosky & Hertzog, 2000). Both of these bases can provide accurate information about the effectiveness of study strategies, but they can also be misleading. Related to the present study, Wahlheim et al. (2012) found that participants were aware of the benefits of variability for classification of novel exemplars on questionnaires that invited consideration of theoretical beliefs in addition to prior experience, but participants were misled by the fluency created by repetitions during study when making predictions of later performance for individual categories. Thus, the effectiveness of participant study preferences should depend on the primary basis for their choices.

Study choices made on the basis of prior experience with category exemplars may reflect the extent to which participants are aware of differences in classification difficulty across categories. Several recent studies have shown that participants are aware of these differences when making predictions of classification performance for novel exemplars. Most relevant to the present study, participants have consistently shown awareness of difficulty differences across categories of birds (i.e., Jacoby, Wahlheim, & Coane, 2010; Tauber & Dunlosky, 2015; Wahlheim et al., 2011, 2012). In addition, participants' awareness of these differences has recently been replicated and extended to other category domains. Rawson, Thomas, and Jacoby (2015) showed that participants were aware of differences across declarative concepts often found in psychology courses, such as the *availability heuristic.* Similarly, Thomas, Finn, and Jacoby (2015) showed that participants were aware of differences across categories of trivia questions that represented different topics that students may learn in educational settings. Together, this collection of studies shows converging evidence that participants' relative accuracy in discriminating among categories is quite reasonable.

Overview of the present experiments

The present experiments examined participant preferences for variability following both recent experience with category exemplars and when recent experience was not provided. Providing experience with exemplars prior to judgements allowed us to examine the relationship between metacognitive judgements made for individual categories and subsequent study choices. In contrast, requiring participants to make choices without recent experience allowed us to examine whether study preferences differed when the primary basis for choices was pre-existing theoretical beliefs. These experiments extended on Wahlheim et al. (2012) in that study preferences for variability and repetitions were examined for naturally occurring categories (i.e., bird families).

In Experiments 1–3, participants were first presented with a subset of exemplars from various categories and were then instructed to make metacognitive assessments of learning for each category. After making their assessments, they were instructed to choose whether to study each of those categories with more variability or more repetitions of exemplars. In Experiment 4, participants were presented with vignettes asking them to imagine having studied birds from various families. Following that, participants were then asked to make study choices for classification of novel exemplars similar to Experiments 1–3, except that these choices were global instead of being made for individual categories. We describe the procedures, possible outcomes, and their theoretical implications in more detail prior to each experiment below.

Experiment 1

Experiment 1 examined the extent to which participants preferred variability when studying and whether prior experiences with exemplars influenced those choices. Participants were first familiarised with exemplars from different categories through intermixed presentation. They were then asked to make predictions of later classification of novel exemplars from each category (via CLJs). In a following phase, participants were asked to choose whether they wanted their additional study to include more variability or repetitions of exemplars. Based on the earlier studies mentioned above showing that participants are aware of difficulty differences across categories, we expected a positive relationship between CLJs and classification of novel exemplars. In addition, based on the finding from Wahlheim et al. (2012) that the majority of participants preferred variability on questionnaire responses, we though it reasonable that the same could be true here. However, it was unclear how participants would make study choices for individual categories.

To gain purchase on this issue, we compared CLJ magnitudes between study options for variability and repetitions for participants who did not have an exclusive preference. In terms of predictions, more difficult categories typically contain fewer overlapping features among exemplars, whereas the opposite is true for less difficult categories. One possibility is that participants will assign categories of greater difficulty to receive more variability to provide additional opportunities to learn the characteristic features common among exemplars. Alternatively, participants without exclusive preferences might instead assign categories of lower difficulty to receive more variability in order to explore category boundaries and form more complete representations of similarity space. Finally, participants may make their choices more generally on the basis of other idiosyncratic beliefs, resulting in no systematic differences in the assignment of categories to specific study options based on assessed difficulty.

Method

Participants

Sixty students from Washington University in St. Louis (18 men, $M_{\text{Age}} = 19.87$ years, Range = 18–22 years) participated for course credit or $10 and were tested individually. The sample size chosen here was exploratory, because research on study choices in category learning is a new area of inquiry. No analyses were conducted until data collection was completed.

Materials

Pictures of perching birds from the taxonomic order Passeriformes (see Figure 1) were taken from an earlier study (Wahlheim et al., 2012). The full set included 12 categories (e.g., bird families such as *Sparrow*), each comprised of 12 exemplars (e.g., bird species such as *chipping sparrow*).

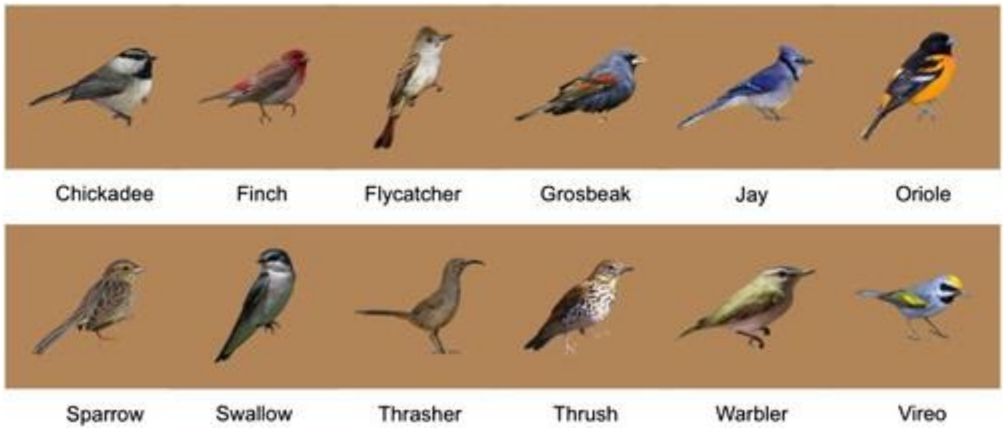

Figure 1. Examples of exemplars from each category.

Procedure and design

Figure 2 displays the eight phases of the experiment: (1) familiarisation, (2) similarity ratings, (3) initial CLJs, (4) study choices, (5) study, (6) updated CLJs, (7) classification of novel exemplars, and (8) classification of studied exemplars.

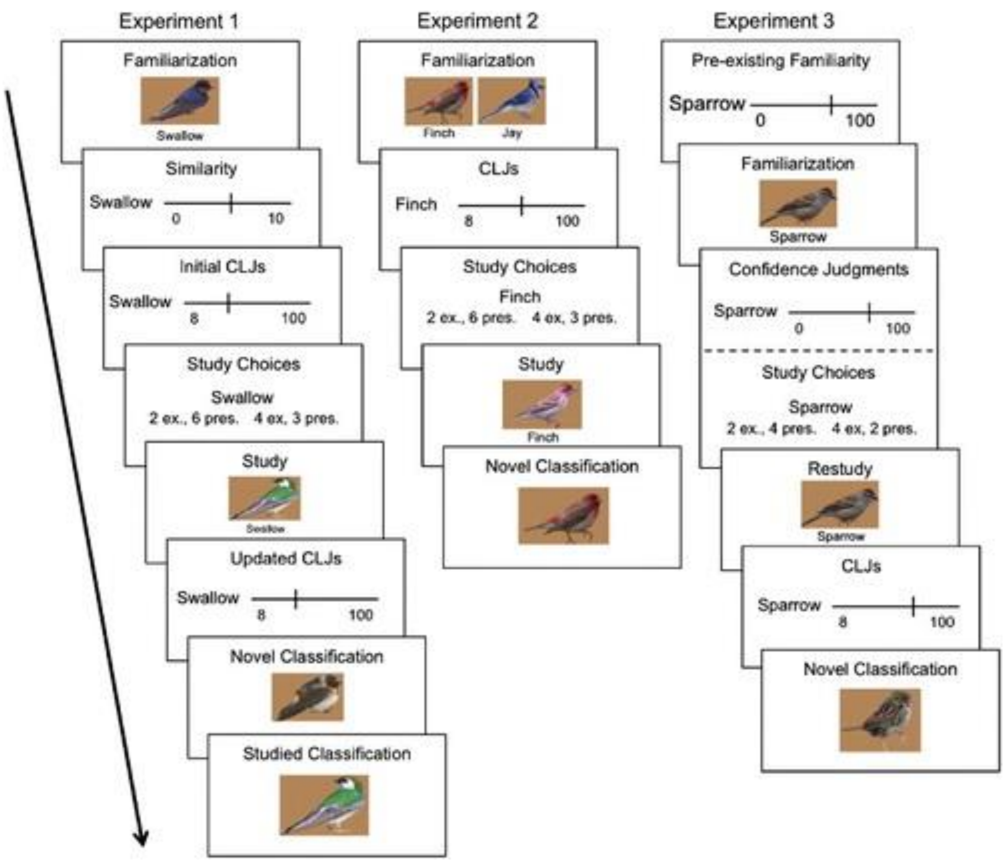

Figure 2. Schematic of the procedures: Experiments 1–3.

During familiarisation, 4 exemplars from each of the 12 categories were selected randomly, for a total of 48. Exemplars were presented individually in random order for 8 s each above their category label, followed by a 500 ms interstimulus interval (ISI). Participants were instructed to study exemplars, read labels aloud, and consider the similarities among exemplars. The 12 category labels then appeared vertically, each next to a sliding scale. Participants rated the extent to which each category was similar to other categories in general on a scale from 0 (*not similar*) to 10 (*very similar*). This was done to highlight the differentiating features among categories. Participants then predicted the probability that they would correctly classify novel exemplars correctly on a subsequent test by making initial CLJs on a scale from 8% (*guessing*) to 100% (*certain correct*). The lower bound was set to approximate chance performance.

Participants were then instructed to structure their study of a set of novel exemplars from categories presented in the familiarisation phase to prepare for a final set of never-before-seen novel exemplars. They were presented with the following text:

In preparation for your test of new species of birds from the earlier studied families, you will be allowed to study additional new species from the families you studied earlier. In this next part, you will be able to decide the manner in which additional birds will be presented to you. For each family, you will be allowed to choose a combination of number of new species and number of presentations from each species. You will have 2 choices that each result in 12 presentations of birds from a family. The combinations will be either 2 new species presented 6 times, or 4 new species presented 3 times each. When making your selection for each family, think about which combination would allow you to best classify a set of new species that will be presented later in the experiment. The 2 options from which to choose will appear beneath each family name.

Participants were then given two options for each category: a variability option (4 exemplars × 3 repetitions) and a repetitions option (2 exemplars × 6 repetitions). Both options produced 12 presentations for each of the 12 categories (144 total). All 12 category labels appeared on the screen simultaneously, each above a drop-down menu that contained both study options. Participants were instructed to choose one option for each of the categories. Novel exemplars then appeared for study in random order in the manner chosen by participants. Each exemplar appeared for 8 s followed by a 500 ms ISI. After study, participants updated their predictions (made a new set of CLJs) using the same scales as before and were instructed to consider the learning accomplished during study when making their judgements.

Finally, participants completed a classification test of novel exemplars followed by test of studied exemplars. In these tests, exemplars appeared individually in random order. Novel classification included four never-before-seen exemplars from each studied category (48 total), and studied classification included 2–4 studied exemplars per category (24–48 total, depending on study choices). For both tests, the 12 category labels appeared to the right of each exemplar, and participants clicked one to classify each exemplar. All participant decisions throughout the experiment were self-paced.

Results and discussion

To assess preference for variability, we classified participants based on the number of categories they assigned to the variability option. In Experiments 1–3, participants who assigned 7–12 categories to the variability option were considered to have a preference for variability during study, participants who assigned 0–5 categories to the variability option were considered to prefer repetitions during study, and participants who assigned six categories to the variability option were considered to prefer neither variability nor repetitions. Table 1 (top row) shows that more participants preferred the variability option (60%) over the repetitions option (23%) and neither option (17%), smallest $\chi^2(1) = 9.68$, $p = .002$. The percentage of participants who preferred the repetitions option did not differ from the percentage who preferred neither option, $\chi^2(1) = 0.67$, $p = .41$. The distribution of individual differences across participants is displayed in Figure 3 (left panel). As the figure shows, a substantial percentage of participants exclusively preferred variability, but a few participants exclusively preferred repetitions. In addition there were many participants who showed non-exclusive preferences for one or the other study option. These individual differences hint that participants may have based their choices on a combination of theoretical beliefs and prior experiences.

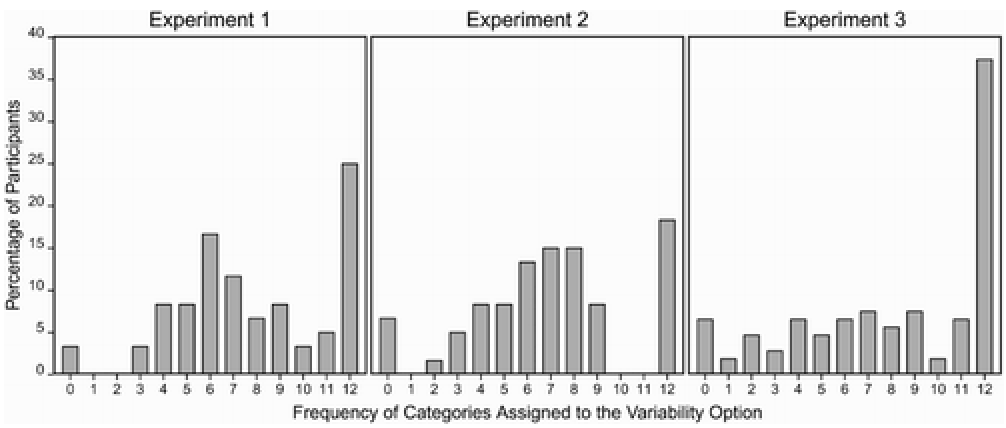

Figure 3. Study choice distributions depicting the percentage of participants who assigned each possible frequency of categories to the variability option: Experiments 1–3.

Table 1. Participant preferences as a function of experiment and familiarization: Experiments 1–3.

	Participant preference			
Experiment	**Familiarisation**	**Variability**	**Repetitions**	**None**
Experiment 1	Intermixed	60%	23%	17%
Experiment 2	Intermixed	63%	27%	10%
	Blocked	50%	33%	17%
Experiment 3	Intermixed	66%	27%	7%
Experiments 1–3		62%	27%	11%

We examined the extent to which experience with prior exemplars influenced study choices by verifying that CLJs reflected difficulty differences across categories. Figure 4 (left panel) displays the relationship between initial CLJs and novel classification performance for each category, averaged across all participants. A Pearson correlation revealed a significantly positive linear relationship between these variables, $r(10) = .84$, $p = .001$. Converging evidence for this relationship was also found, as the mean within-participant correlation ($M = .25$, SD = .30) was also significantly greater than zero, $t(59) = 6.45$, $p < .001$, $d = .83$. Together, these correlations show that participants were aware of difficulty differences among categories.

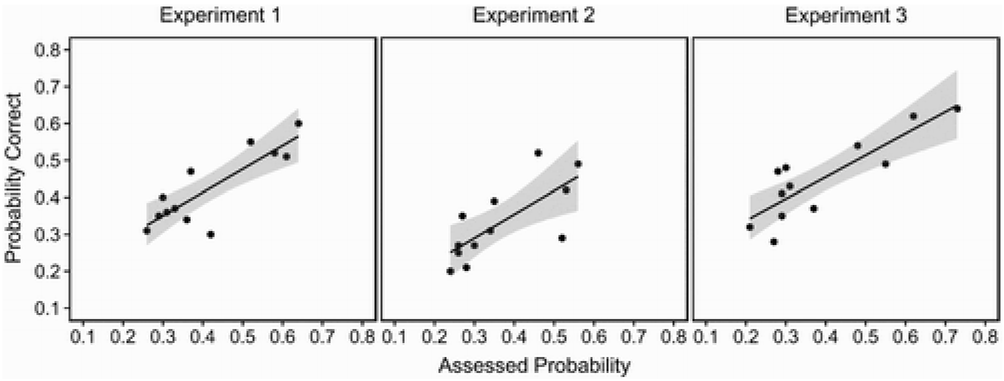

Figure 4. Scatterplots depicting the group-level relationships between assessed probabilities of category learning and classification performance on novel exemplars from each category. The metacognitive judgements made prior to study choices were CLJs (Experiments 1 and 2) and confidence judgements (Experiment 3). Error bands represent 95% confidence intervals.

We next compared CLJ magnitudes (both initial and updated) for categories assigned to the variability and repetitions options for participants who did not have exclusive preferences (an approach we repeat in Experiments 2 and 3). Despite participants' awareness of difficulty differences, study choices were not related to initial CLJs. Table 2 (top row) shows that CLJ magnitudes did not differ between the variability ($M = 44.00$, SD = 18.24) and repetitions ($M = 42.54$, SD = 21.94) study options, $t(42) = 0.31$, $p = .80$, $d = .05$. This result suggested that participants did not systematically assign categories to study options based on perceived difficulty of classification. Although study choices were not based on initial CLJs, participants' updated CLJs suggested that they indeed learned from the study phase. Judgement magnitudes were greater for updated CLJs ($M = 55.19$, SD = 12.97) than initial CLJs ($M = 43.27$, SD = 12.89), $F(1, 42) = 48.66$, $p < .001$, $\eta_p^2 = .54$. However, after the study phase, participants did not predict that having had more variability among studied exemplars would benefit later classification of novel exemplars beyond having had more repetitions, as updated CLJs did not differ between categories assigned to the variability ($M = 54.16$, SD = 14.91) and repetitions ($M = 56.23$, SD = 19.08) study options, $t(42) = 0.61$, $p = .55$, $d = .09$.

Table 2. Metacognitive judgement magnitudes as a function of judgement type and study choice: Experiments 1–3.

			Study Choice	
Experiment	Familiarisation	Judgement Type	Variability	Repetitions
Experiment 1	Intermixed	Initial CLJs	44.00 (5.61)	42.54 (6.75)
		Updated CLJs	54.16 (4.59)	56.23 (5.87)

Experiment 2	Intermixed	CLJs	36.93 (8.49)	42.57 (8.84)
	Blocked	CLJs	31.39 (8.68)	34.90 (9.03)
Experiment 3	Intermixed	Confidence	47.28 (6.37)	51.04 (6.67)
	Intermixed	Familiarity	19.31 (4.82)	24.24 (5.70)
	Intermixed (Honoured)	CLJs	56.21 (7.75)	66.75 (8.79)
	Intermixed (Dishonoured)	CLJs	55.69 (7.24)	56.09 (8.22)

Finally, classification performance for participants without exclusive preferences replicated earlier studies showing that performance for novel exemplars was significantly higher for the variability ($M = .48$, SD = .15) than repetitions ($M = .39$, SD = .17) study option, $t(42) = 3.15$, $p = .003$, $d = .48$, whereas there was a non-significant trend showing that performance for studied exemplars was numerically greater for the repetitions ($M = .74$, SD = .23) than variability ($M = .69$, SD = .17) study option, $t(42) = 1.44$, $p = .16$, $d = 0.23$. The Item Type × Study Option interaction was significant, $F(1, 42) = 20.03$, $p < .001$, $\eta_p^2 = .32$. Taken with the finding that updated CLJs did not differ between categories assigned to the variability and repetitions study options, these results replicated earlier studies showing variability neglect (Doyle & Hourihan, 2015; Wahlheim et al., 2012).

Experiment 2

Experiment 1 showed that most participants preferred variability during study but that there were individual differences in the extent to which either strategy was preferred. In addition, participants who did not have an exclusive study preference did not appear to make their study choices on the basis of prior assessments of difficulty differences across categories. The primary aims of Experiment 2 were to determine if participant preferences for variability would replicate and to further examine whether those preferences reflect prior experience with exemplars. The procedure in Experiment 2 was similar to that used in Experiment 1 in that participants were given experience with exemplars prior to making CLJs and subsequent study choices about variability and repetitions. To determine if prior experience influences study choices, we manipulated the presentation order of exemplars during the familiarisation phase.

One group of participants studied pairs of exemplars from different categories and made judgements about the extent to which each pair had different features (Intermixed group). Another group of participants studied pairs of exemplars from the same categories and made judgements about the extent to which each pair had similar features (Blocked group). Studies comparing intermixed and blocked presentation of exemplars have suggested that intermixed study encourages noticing of differences, whereas blocked study encourages noticing of similarities (Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Kornell & Bjork, 2008; Rohrer, Dedrick, & Stershic, 2015; Wahlheim et al., 2011). Varying the presentation of exemplars in this way during familiarisation could encourage participants in each group to examine exemplars with different goals. One possibility is that more participants will prefer variability in the Intermixed than Block group because the former encourages participants to search for differences while the latter encourages participants to search for similarities. However, another possibility is that theoretical beliefs will still serve as the dominant basis for study choices, resulting in differences in prior experience having no effect on the percentages of participant preferences.

Method

Participants

Sixty students from Washington University in St. Louis participated in exchange for course credit or $10 and were tested individually. Participants were randomly assigned in even numbers to the Blocked ($N = 30$, 14 men, $M_{\text{Age}} = 19.00$, Range = 18–21 years) and Intermixed ($N = 30$, 14 men, $M_{\text{Age}} = 18.97$, range = 18–21 years) groups. This sample size was chosen because a power analysis for a chi-squared test with two degrees of freedom with alpha = .05, power = .80, and an estimated medium effect size indicated that a total sample of 60 participants would be sufficient to detect an effect. No analyses were conducted until after data collection was completed.

Procedure and design

Figure 2 displays the five phases of this experiment: (1) familiarisation, (2) CLJs, (3) study choices, (4) study, and (5) classification of novel exemplars.

The familiarisation phase presented four exemplars each from 12 categories (48 total). Exemplars appeared in pairs for 16 s each (24 total trials). Each pair in the Intermixed group included exemplars from different categories. This was achieved by taking one exemplar from each category at random, randomising the order of those 12, presenting two at a time, and then repeating the process three more times. This meant that one exemplar from each category was presented before the next set of exemplars from each category. In the Blocked group, each pair included exemplars from the same categories for two consecutive presentations. After each pair of exemplars from different categories was presented, participants rated differences between exemplars on a sliding scale from 0 (*not different*) to 10 (*extremely different*). After pairs of exemplars from the same categories, participants rated similarities between exemplars on a sliding scale from 0 (*not similar*) to 10 (*extremely similar*). The remaining phases proceeded as in Experiment 1. All participant decisions throughout the experiment were self-paced.

Results and discussion

The percentages of participant preferences presented in Table 1 (middle rows) did not differ as a function of presentation order during familiarisation, $\chi^2(2, N = 60) = 1.19$, $p = .55$, $\varphi = .14$. However, the pattern of participant preferences largely replicated results from Experiment 1, as the overall percentage of participants who preferred variability (57%) was significantly greater than the percentage of participants who preferred repetitions (30%), $\chi^2(1) = 4.92$, $p = .03$, and the percentage of participants who preferred repetitions was marginally greater than the percentage of participants who preferred neither option (13%), $\chi^2(1) = 3.85$, $p = .05$. Results from a two-sample Kolmogorov-Smirnov test revealed no significant difference between the distributions (collapsed in Figure 3, middle panel), $D = .78$, $p = .59$, providing further evidence that the presentation order during familiarisation did not influence choices.

Participants' CLJs again indicated awareness of difficulty differences across categories. The group-level Pearson correlations between CLJs and novel classification performance for all participants did not differ between the Intermixed, $r(10) = .78$, and Blocked, $r(10) = .63$, groups, $z = 0.64$, $p = .52$. Consequently, the correlation was computed for the entire sample, $r(10) = .72$, $p = .008$, and the collapsed data are presented in Figure 4 (middle panel). The mean within-participant Pearson correlations between CLJs and novel classification performance also did not differ between the Intermixed ($M = .28$, SD = .31) and Blocked ($M = .19$, SD = .30) groups, $t(57) = 1.19$, $p = .24$, $d = .30$, and the correlation for the entire sample ($M = .24$, SD = .31) was significantly greater than zero, $t(58) = 5.89$, $p < .001$, $d = .78$ (the correlation could not be computed for one participant due to constant confidence values).

The relationships between study choices, CLJs, and classification performance were examined only for participants who did not have an exclusive study preference. Despite participants' awareness of difficulty differences across categories, CLJs did not differ between categories assigned to the variability ($M = 34.22$, SD = 20.15) and repetitions ($M = 38.82$, SD = 21.13) study options, $F(1, 43) = 0.96$, $p = .33$, $\eta_p^2 = .02$. CLJ magnitudes also did not differ between the Intermixed ($M = 39.75$, SD = 11.55) and Blocked ($M = 33.14$, SD = 15.05) groups, $F(1, 43) = 2.74$, $p = .11$, $\eta_p^2 = .06$. There was not a significant Familiarisation × Study Option interaction, $F(1, 43) = 0.05$, $p = .82$, $\eta_p^2 < .01$. Classification performance on novel exemplars did not differ between categories assigned to the variability ($M = .39$, SD = .14) and repetitions ($M = .34$, SD = .17) options, nor between the Intermixed ($M = .39$, SD = .10) and Blocked ($M = .34$, SD = .10) groups, *largest* $F(1, 43) = 1.81$, $p = .19$, $\eta_p^2 = .04$. The Study × Group interaction was not significant, $F(1, 43) = 0.75$, $p = .39$, $\eta_p^2 = .02$. This failure to replicate the benefits of variability shown in Experiment 1 may have resulted from differences in encoding operations created by the manipulation of presentation order during familiarisation.

Experiment 3

Experiments 1 and 2 showed that participants were aware of difficulty differences across categories, most participants preferred variability in their study, and study choices were not related to CLJs for participants without exclusive preferences. One possible reason that CLJs did not differ between study options is that CLJs were made for all categories in the phase prior to study choices. This may have given participants more time to consider other bases for their decisions. Experiment 3 examined whether the patterns in Experiments 1 and 2 showing no differences in metacognitive judgements at the category level between study options would still persist when participants were instructed to retrospectively assess their learning of categories from the familiarisation phase immediately before making a study choice for each. Finding that study choices are not related to metacognitive judgements under these conditions would provide more convincing support for the idea that other bases, such as theoretical beliefs, play a dominant role in determining study preferences for variability.

Another aim of Experiment 3 was to determine whether participant study choices were adaptive for later classification of novel exemplars. It is possible that individual differences in study choices reflect accurate beliefs about which of the two study strategies will be most effective for individual categories, given the conditions of the experiment. To examine whether participants made adaptive choices, we employed the honour/dishonour procedure introduced by Kornell and Metcalfe (2006). In our variant of this procedure, participants chose whether to restudy exemplars presented during familiarisation with more variability or more repetitions. During restudy, the honoured group was presented with exemplars in the manner that they requested, whereas the dishonoured group was presented with exemplars in the manner opposite of what they requested. Evidence that participants made adaptive choices would be shown by greater classification of novel exemplars in the Honoured than Dishonoured group.

Method

Participants

One hundred seven participants from Amazon Mechanical Turk participated at their own computers in exchange for $1. Participants were randomly assigned to either the Honoured ($N = 56$; 33 men, $M_{\text{Age}} = 37.32$, Range = 18–67 years) or Dishonoured ($N = 51$; 28 men, $M_{\text{Age}} = 37.04$, Range = 18–67 years) group. Given that we did not find differences between CLJs for categories assigned to variability and repetitions options in Experiments 1 and 2, and that no previous experiments have established an effect size for differences in classification performance for these materials between honoured and dishonoured groups, we employed the rule that data collection would cease when the experiment had been posted for at least one week and there were at least 30 participants per group (comparable sample as in Experiment 2). Analyses were performed only after data collection was completed.

Procedure and design

Figure 2 displays the six phases of the experiment: (1) familiarity ratings, (2) familiarisation, (3) confidence judgements and study choices, (4) restudy, (5) CLJs, and (6) classification of novel exemplars.

Prior to the experiment, participants were instructed to rate their pre-existing familiarity with each of the categories. This measure was included to allow for an exploratory analysis of the extent to which prior knowledge influences study choices. Category labels appeared individually in random order, and participants rated the familiarity of each on an adjacent scale ranging from: 0 (*not at all familiar*) to 10 (*completely familiar*).

Once the experiment began, participants received an intermixed familiarisation phase as in earlier experiments. Exemplars appeared for 5 s each, and participants were required to press the space bar to advance during ISIs to encourage their sustained attention to the task. Next, category labels appeared individually in random order, and participants made two judgements for each. Participants first rated how confident they were that they learned the exemplars during the familiarisation phase on a scale from: 0% (*not at all learned*) to 100% (*completely learned*). Participants then chose how they would like to *restudy* exemplars from the familiarisation phase for an upcoming test of novel exemplars. The choice to restudy exemplars (instead of study new ones as in Experiments 1 and 2) was employed to more closely parallel the original honour/dishonour procedure (cf. Kornell & Metcalfe, 2006). Participants could choose between the variability option, which included 4 exemplars × 2 repetitions, and the repetitions option, which included 2 exemplars × 4 repetitions. Both groups were warned that their choices might not be honoured during study. Study choices included fewer presentations than in earlier experiments to shorten the procedure for online administration. The instructions to this phase read:

You will choose how to restudy birds from each family. One option will be to restudy all of the birds you saw in the previous phase, with each bird being presented twice (4 birds, 2 presentations each). The other option will be to restudy only two of the birds you saw in the previous phase, with each bird being presented four times (2 birds, 4 presentations each). Again, you should choose the method of study that you believe will optimize your ability to classify new birds from these families on a later test.

In the study phase, the honoured group received their study choices, whereas the dishonoured group received the opposite of their choices. Participants pressed the space bar to advance during ISIs. The remaining phases followed as in Experiments 1 and 2, except that CLJs made after the study phase were made for category labels that appeared individually and in random order. All participant decisions throughout the experiment were self-paced.

Results and discussion

The overall percentages of participant preferences for variability, repetitions, or neither option were computed for the entire sample collapsed across groups, because the honour/dishonour manipulation did not occur until after study choices were made. Similar to Experiments 1 and 2, Table 1 shows that the percentage of participants who preferred variability (66%) was significantly greater than the percentage of participants who preferred repetitions (27%), and the percentage of participants who preferred repetitions was significantly greater than the percentage of participants who preferred neither option (7%), *smallest* $\chi^2(1) = 13.44$, $p < .001$. Figure 3 (right panel) shows that most participants had an exclusive preference for variability, but there was a reasonable range of preferences across all possible combinations.

As with CLJs in Experiments 1 and 2, confidence judgements reflected awareness of difficulty differences across categories. Figure 4 (right panel) displays the relationship between confidence judgements and novel classification performance for individual categories averaged across all participants. Replicating Experiments 1 and 2, the Pearson correlation collapsed across all participants indicated that there was a significantly positive linear relationship between these variables, $r(10) = .86$, $p < .001$. The mean within-participant correlation was also positive ($M = .29$, SD = .29) and significantly greater than zero, $t(105) = 10.46$, $p < .001$, $d = 1.00$ (the correlation could not be computed for one participant due to constant confidence values).

The relationships between study choices, pre-existing familiarity, confidence judgements, CLJs, and classification performance were examined only for participants who did not have an exclusive study preference. Table 2 (fifth row) shows that although confidence judgements reflected participants' awareness of classification difficulty across categories, study choices were not related to confidence judgements. The magnitude of confidence judgements did not differ between categories assigned to the variability ($M = 47.28$, SD = 24.41) and repetitions

($M = 51.04$, SD = 26.16) study options, $t(59) = .91$, $p = .37$, $d = 0.12$. Together, these results show that requiring participants to retrospectively consider differences in learning across categories and to make study choices immediately following such reflection did not elicit a relationship between assessments of category learning and study choices.

Table 2 also shows that study choices were not made on the basis of pre-existing familiarity with individual categories in that familiarity judgements did not differ between categories assigned to variability ($M = 19.31$, SD = 18.55) and repetitions ($M = 24.24$, SD = 21.82) study options, $t(59) = 1.58$, $p = .12$, $d = 0.21$.

Finally, Table 2 shows that CLJs made after the study phase did not differ between the variability ($M = 55.93$, SD = 20.30) and repetitions ($M = 61.07$, SD = 22.66) options, $F(1, 58) = 2.62$, $p = .11$, $\eta_p^2 = .04$. The effect of Group, and the Study × Group interaction were not significant, *largest* $F(1, 58) = 2.25$, $p = .14$, $\eta_p^2 = .04$. These results are consistent with earlier findings showing that participants did not predict benefits of variability for classification of novel exemplars. However, these predictions were consistent with actual performance in that classification performance for novel exemplars did not differ between the variability ($M = .37$, SD = .20) and repetitions ($M = .40$, SD = .19) study options, $F(1, 58) = 0.74$, $p = .39$, $\eta_p^2 = .01$. This lack of a variability benefit was likely due to the fact that restudying a subset of exemplars from the familiarisation phase resulted in fewer unique exemplars being presented during study, relative to earlier experiments. This had consequences for whether participants could benefit from control over their study choices; classification of novel exemplars did not differ between the honour ($M = .37$, SD = .15) and dishonour ($M = .40$, SD = .14) groups, $F(1, 58) = 1.05$, $p = .31$, $\eta_p^2 = .02$, even though participants generally preferred variability during study. The Study × Group interaction was not significant, $F(1, 58) = 0.02$, $p = .89$, $\eta_p^2 < .01$. This methodological feature rendered the interpretation of honour/dishonour results somewhat ambiguous. Nonetheless, Experiment 3 added value in showing study preferences for variability consistent with Experiments 1 and 2.

Experiment 4

Experiments 1–3 showed that neither predictions nor postdictions of learning for individual categories were related to study choices, despite the fact that participants were aware of difficulty differences across categories. These results were interpreted as showing that participants' choices were based more heavily on a basis other than prior experience, such as theoretical beliefs about the effectiveness of study strategies. However, participants may not have considered complex theories about study effectiveness but instead may have preferred variability because they thought that different pictures would be more interesting to look at. Experiment 4 was designed to illuminate these possibilities by requiring participants to make study choices without having experienced the actual materials (cf. McCabe, 2011).

In Experiment 4, participants read vignettes describing study and test scenarios akin to those given to participants in Experiments 1–3. Participants were then instructed to make global judgements about whether more variability or more repetitions would be a better strategy for later classification performance. One group of participants made these judgements for classification of both studied and novel exemplars on the same sheet of paper, similar to the questionnaires given to participants in Wahlheim et al. (2012). Another group of participants only made these judgements for classification of novel exemplars. If participants simply prefer variability in their study because it provides them with a more interesting viewing experience, then the majority of participants should indicate that variability is more effective than repetitions for both studied and novel exemplars. Alternatively, if participants make their choices based on theories about how these variables affect learning and later classification performance, then study choices for novel exemplars should show a preference for variability similar to Experiments 1–3, whereas choices for studied exemplars should show a preference for repetitions as in Wahlheim et al. (2012).

Finally, the extent to which participants indicate that variability is a more effective study strategy for classification of novel exemplars might depend on whether judgements are made for both studied and novel exemplars, or only for novel exemplars. Studies from the decision-making literature have shown that manipulations of choice architecture accomplished by varying the configuration of response options can dramatically shift decisions (see Thaler & Sunstein, 2008). Thus, more participants might indicate a preference for variability when making judgements for both types of exemplars simultaneously, because doing so encourages reliance on analytic bases for decisions.

Method

Participants

One hundred students from Washington University participated in exchange for course credit or $10 per hour (36 men, mean age = 19.3). Fifty students were assigned randomly to each of the questionnaire groups and were tested individually. The sample sizes here were selected by performing a power analysis that determined the number of participants necessary to achieve a medium effect-size using a chi-squared test with one degree of freedom, power set to .80, and alpha set to .05. This analysis indicated that 100 participants would be more than sufficient.

Design, materials, and procedure

Participants were given a questionnaire that asked them to imagine that they would have the opportunity to study a number of different bird species (e.g., chipping sparrow) belonging to different families (e.g., Sparrow). In a Studied and Novel group, participants were asked how they would study for a classification test of bird species that they had seen before (i.e., studied exemplars), and were then asked the same question about novel exemplars. Both questions appeared on the same piece of paper to encourage participants to contrast the questions for each type of test item. The question about studied exemplars always appeared above the question about novel exemplars, but participants were allowed to change their responses. In contrast, in the Novel Only group, participants were asked to choose how they would study if the goal was only to classify bird species that they had never seen before (i.e., novel exemplars). For both groups, participants chose between a *more variability* option that included four species from each family presented three times during study and a *more repetitions* option that included two species from each family presented six times. The exact vignettes are displayed in Table 3. Note that these two options were equated for number of presentations in that both study choices afforded participants 12 presentations of exemplars. The arrangement of study choices within each question on both questionnaires was counterbalanced. Finally, participants were instructed to write a few sentences to justify their choices. This feature was useful for determining whether participants who preferred variability in their choices about novel exemplars did so simply because more pictures would be interesting to look at.

Table 3. Vignettes presented in the studied and novel and novel only groups: Experiment 4.

<table>
<tr><td colspan="2">Studied and novel
Imagine that you were tasked with studying pictures of various species of birds belonging to several families along with the name of the family to which they belong (e.g., a chipping sparrow with the family name sparrow). In this situation you would have a limited number of trials for study, but you would be able to choose the number of species you could study and the number of times you could repeat the study of each species. However, because of the limited trials, there would be a trade-off between the number of species and the number of repetitions. If your goal was to maximise the accuracy with which you could classify studied species (e.g., the studied chipping sparrow) on a test 5 minutes later, which would you prefer to study?</td></tr>
<tr><td>More species than repetitions
(e.g., four species, three repetitions)</td><td>Fewer species than repetitions
(e.g., two species, six repetitions)</td></tr>
<tr><td colspan="2">Now imagine the same scenario, but instead of preparing for a later classification test of studied species, you would be preparing for a test of unstudied species belonging to studied families (e.g., classifying a never-before-seen tree sparrow as a member of the sparrow family). If your goal was to maximise the accuracy with which you could classify unstudied species belonging to those families (e.g., the unstudied tree sparrow) on a test 5 minutes later, which would you prefer to study?</td></tr>
<tr><td>More species than repetitions
(e.g., four species, three repetitions)</td><td>Fewer species than repetitions
(e.g., two species, six repetitions)</td></tr>
<tr><td colspan="2">Novel Only
Imagine that you were tasked with studying pictures of various species of birds belonging to several families along with the name of the family to which they belong (e.g., a chipping sparrow with the family name Sparrow). In this situation, your goal would be to learn birds so as to maximise the accuracy with which you could later classify new species belonging to those families (e.g., classifying a never-before-seen tree sparrow as a member of the sparrow family on a test 5 minutes later). In addition, you would have a limited number of trials for study, but you would be able to choose the number of species you could study and the number of times you could repeat the study of each species. However, because of the limited trials, there would be a trade-off between the number of species and the number of repetitions. With the goal of later classification of unstudied species, which would you prefer to study?</td></tr>
<tr><td>More species than repetitions
(e.g., four species, three repetitions)</td><td>Fewer species than repetitions
(e.g., two species, six repetitions)</td></tr>
</table>

Results and discussion

Table 4 (top row) shows that nearly every participant in the Studied and Novel group predicted that classification of studied exemplars would be better following more repetitions (92%), whereas only a few participants predicted that it would be more beneficial to receive additional variability (8%), $\chi^2(1) = 35.28$, $p < .001$. In contrast, when those same participants made their choices for classification of novel exemplars, more participants preferred variability (66%) than repetitions (34%), $\chi^2(1) = 5.12$, $p = .02$. Finally, changing the question structure such that participants were only asked to make choices about which option would benefit later classification of novel exemplars produced the opposite pattern. Fewer participants in the Novel Only group preferred variability (30%) than repetitions (70%), $\chi^2(1) = 8.00$, $p = .005$.

Table 4. Participant beliefs about the effectiveness of study strategies for classification of novel exemplars as a function of question structure and exemplar type: Experiment 4.

		Study choice	
Question structure	Exemplar type	Variability (%)	Repetitions (%)
Studied and novel	Studied	8	92
	Novel	66	34
Novel only	Novel	30	70

The finding of a preference reversal for classification of novel exemplars depending on the context in which the question was presented suggests that more participants preferred variability when they had previously considered classification of studied exemplars. Perhaps the contrast between the types of test items motivated participants to think more analytically about how the effects of variability and repetitions encoding methods would differ. Moreover, this reversal cannot be fully explained by an account holding that participants prefer more variability because new pictures are more interesting to look at. Consistent with this conclusion, only one participant from the Novel Only condition indicated preferring variability for a test of novel exemplars because more exemplars would enhance the viewing experience. Specifically, the participant reported, “I get uninterested if there are too many repetitions. A change of pace keeps things interesting.” Thus, although it is possible that one reason participants from the earlier experiments assigned the majority of categories to the variability option was to preclude boredom during study, it is unlikely that most participants preferred variability solely for that reason.

General discussion

The present experiments showed that the majority of participants preferred variability when studying exemplars from naturally occurring categories for a later test of novel exemplars from those categories. Participants showed individual differences in their preferences that were not related to category-level metacognitive assessments of learning made after initial exposure to representative exemplars (Experiments 1–3). These findings suggested that participants made their choices primarily on the basis of theoretical beliefs. Consistent with this interpretation, most participants preferred variability when making choices in the absence of experience with category exemplars, but only when they considered categorisation of both studied and novel exemplars (Experiment 4).

Although participants were likely to have made theory-based choices in the present experiments, it is possible that participants differed in the theories that they employed. One adaptive approach to study choices would have been to consider reasons why one strategy may be more effective than another, and to select the more effective strategy. Given that the majority of participants preferred variability in their study, with many of those participants reporting an exclusive preference for variability (Experiments 1–3), it is possible most participants were aware that variability often benefits novel classification. It is also possible that participants, to some extent, preferred variability during study because new pictures would be more interesting to look at. The results from the questionnaire data in Experiment 4 suggest that this basis did contribute to at least one participant's judgements, but it may not have been the primary reason why all participants preferred variability. However, participants in Experiments 1–3 may have been more likely to prefer variability due to the increase in interest after having seen several presentations during familiarisation, especially in Experiment 3 where study choices involved restudying exemplars from the familiarisation phase. These findings point to the need for future studies to directly explore how multiple bases jointly contribute to overall and category-level justifications for study choices.

In this vein, one possibility is that recent experiences inform theory-based choices by encouraging more analytic thinking about which strategy would more effective, regardless of the details of individual categories. Preliminary evidence for this possibility was shown here in that the majority of participants preferred variability in Experiments 1–3 wherein study choices were preceded by experience with exemplars and in Experiment 4 when participants made choices for both studied and novel exemplars, whereas the majority of participants preferred repetitions in Experiment 4 when they were only asked to make choices for novel exemplars. In addition, it is still possible that the presentation order during familiarisation could influence beliefs about effective study strategies, but the between-subjects manipulation of presentation order in Experiment 2 might have been too subtle to shift choices, despite the fact that each group processed exemplars differently. Perhaps a within-subjects manipulation of presentation order or the number of repetitions would create differences in the processing fluency of prior experiences that are sufficiently salient to alter theoretical notions about the effectiveness of each strategy.

More generally, the nascent enterprise of self-regulated category learning has theoretical implications for understanding how participants approach their learning in everyday contexts. Current frameworks holding that controlled behaviours are directly influenced by metacognitive evaluations of prior learning (e.g., Koriat et al., 2006; Nelson & Narens, 1990) may inform predictions about self-regulated concept learning, but they do not currently possess the level of specificity required to account for results such as those obtained in the present experiments. We propose that a comprehensive theory of study preferences for exemplar variability should account for participants' ability to comprehend and predict interactions among variables. The effects of exemplar variability on classification performance depend, in part, on whether studied or novel exemplars are tested (e.g., Wahlheim et al., 2012) and the similarity structure of categories (e.g., Homa & Cultice, 1984). Consequently, study choices aimed at maximising later performance will be adaptive when participants can properly assess how the relevant variables contribute to learning outcomes.

Consistent with this notion, one participant from the Novel only group in Experiment 4 indicated a preference for variability, but also pointed to the role of category similarity in the efficacy of that preference. The participant reported,

I feel that even though I would only have 3 repetitions, I would still be able to identify some common features between the three. This way, I may get a decent idea of what to look for in 4 species instead of 2. However, it would depend on how similar the families appear.

Finally, a theory of these study preferences must also consider individual differences in the extent to which participants try to learn rules or memorise exemplar-label associations (cf. Little & McDaniel, 2015; Wahlheim, McDaniel, & Little, in press). These tendencies may explain, in part, individual differences in the assignment of categories to study options.

Establishing a theoretical framework of this sort will also allow for predictions in applied settings. In education, for example, students are faced with choices about how to regulate their study across different courses as well as across topics within those courses. Both courses and topics can be considered categories, and it is important that students find optimal ways to arrange their study of book chapters and example problems. It is also important for students to distinguish between strategies that are effective for rote learning and strategies that are effective for category learning to be able to apply conceptual knowledge successfully in novel situations. For example, effective strategies for learning new foreign vocabulary words may not be as effective in preparing for a calculus test that includes previously unpracticed problems. The ability for students to modulate their study strategies across courses and topics depends not only on feedback about performance, but also on feedback about strategy choices.

The example of education as a domain in which self-regulated category learning can be applied also points to potential directions for future research. In addition to the present experiments, several recent studies have examined the learning of naturally occurring perceptual categories (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kirchoff, Delaney, Horton, & Dellinger-Johnston, 2014; Kornell & Bjork, 2008; Noh, Yan, Vendetti, Castel, & Bjork, 2014). A unique feature of materials like these is that the similarity structures across categories vary across domains. This is akin to how some courses in educational settings have exemplars that can be classified on the basis of defining features (e.g., math problems), whereas exemplars in other courses are less well defined and can be classified based on characteristic features (e.g., psychiatric diagnoses). Future studies could systematically investigate how study choices differ across category domains that vary in similarity space such as math problems (Rohrer et al., 2015) and declarative concepts from the social sciences (e.g., Rawson et al., 2015). Consideration of the variety of factors that influence study preferences in rich learning contexts such as these holds promise for revealing the nuances underlying self-regulated category learning.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

1. Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, 41, 392–402. http://dx.doi.org/10.3758/s13421-012-0272-7
2. Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444. http://dx.doi.org/10.1146/annurev-psych-113011-143823doi: 10.1146/annurev-psych-113011-143823
3. Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, 42, 481–495. http://dx.doi.org/10.3758/s13421-013-0371-0 doi: 10.3758/s13421-013-0371-0
4. Doyle, M. E., & Hourihan, K. L. (2015). Metacognitive monitoring during category learning: How success affects future behaviour. *Memory*. http://dx.doi.org/10.1080/09658211.2015.1086805
5. Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging*, 13, 597–607. http://dx.doi.org/10.1037/0882-7974.13.4.597doi: 10.1037/0882-7974.13.4.597
6. Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, 15, 462–474. http://dx.doi.org/10.1037/0882-7974.15.3.462 doi: 10.1037/0882-7974.15.3.462
7. Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 83–94.
8. Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1441–1451. http://dx.doi.org/10.1037/a0020636
9. Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, 26, 97–103. doi: 10.1002/acp.1801
10. Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138, 469–486. http://dx.doi.org/10.1037/a0017341 doi: 10.1037/a0017341

11. Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, 35, 157–175. doi: 10.1006/jmla.1996.0009

12. Kirchoff, B. K., Delaney, P. F., Horton, M., & Dellinger-Johnston, R. (2014). Optimizing learning of scientific category knowledge in the classroom: The case of plant identification. *CBE-Life Sciences Education*, 13, 425–436.

13. Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. http://dx.doi.org/10.1037/0096-3445.126.4.349 doi: 10.1037/0096-3445.126.4.349

14. Koriat, A., Ma'ayan, H., Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135, 36–69. http://dx.doi.org/10.1037/0096-3445.135.1.36 doi: 10.1037/0096-3445.135.1.36

15. Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19, 585–592. http://dx.doi.org/10.1111/j.1467-9280.2008.02127.xdoi: 10.1111/j.1467-9280.2008.02127.x

16. Kornell, N., & Finn, B. (in press). Self-regulated learning: An overview of theory and data. In J. Dunlosky& S. Tauber (Eds.), *The Oxford handbook of metamemory*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199336746.013.23

17. Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 609–622. http://dx.doi.org/10.1037/0278-7393.32.3.609

18. Kornell, N., & Son, L. K. (2009). Participants’ choices and beliefs about self-testing. *Memory*, 17, 493–501. http://dx.doi.org/10.1080/09658210902832915 doi: 10.1080/09658210902832915

19. Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition*, 43, 283–297. http://dx.doi.org/10.3758/s13421-014-0475-1 doi: 10.3758/s13421-014-0475-1

20. McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, 39, 462–476. doi: 10.3758/s13421-010-0035-2

21. Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, 132, 530–542. http://dx.doi.org/10.1037/0096-3445.132.4.530 doi: 10.1037/0096-3445.132.4.530

22. Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463–477. http://dx.doi.org/10.1016/j.jml.2004.12.001 doi: 10.1016/j.jml.2004.12.001

23. Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–173. doi: 10.1016/S0079-7421(08)60053-5

24. Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multilevel induction of categories: Venomous snakes hijack the learning of lower category levels. *Psychological Science*, 25, 1592–1599. http://dx.doi.org/10.1177/0956797614535938 doi: 10.1177/0956797614535938

25. Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363. http://dx.doi.org/10.1037/h0025953 doi: 10.1037/h0025953

26. Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, 27, 483–504. http://dx.doi.org/10.1007/s10648-014-9273-3 doi: 10.1007/s10648-014-9273-3

27. Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625. doi: 10.1037/a0013684

28. Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, 24, 355–367. http://dx.doi.org/10.1007/s10648-012-9201-3 doi: 10.1007/s10648-012-9201-3

29. Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, 107, 900–908. http://dx.doi.org/10.1037/edu0000001 doi: 10.1037/edu0000001

30. Son, L. K., & Kornell, N. (2009). Simultaneous decisions at study: Time allocation, ordering, and spacing. *Metacognition and Learning*, 4, 237–248. http://dx.doi.org/10.1007/s11409-009-9049-1 doi: 10.1007/s11409-009-9049-1

31. Tauber, S. K., & Dunlosky, J. (2015). Monitoring of learning at the category level when learning a natural concept: Will task experience improve its resolution? *Acta Psychologica*, 155, 8–18. http://dx.doi.org/10.1016/j.actpsy.2014.11.011 doi: 10.1016/j.actpsy.2014.11.011

32. Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, 20, 356–363. http://dx.doi.org/10.3758/s13423-012-0319-6 doi: 10.3758/s13423-012-0319-6

33. Thaler, R. H., & Sunstein, C. R. (2008). *Nudge*. London: Penguin Books.

34. Thomas, R. C., Finn, B., & Jacoby, L. L. (2015). Prior experience shapes metacognitive judgments at the category level: The role of testing and category difficulty. *Metacognition and Learning*. http://dx.doi.org/10.1007/s11409-015-9144-4

35. 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, 750–763. http://dx.doi.org/10.3758/s13421-010-0063-y doi: 10.3758/s13421-010-0063-y

36. 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, 703–716. http://dx.doi.org/10.3758/s13421-011-0180-2 doi: 10.3758/s13421-011-0180-2

37. Wahlheim, C. N., McDaniel, M. A., Little, J. L. (in press). Category learning strategies in younger and older adults: Rule abstraction and memorization. *Psychology & Aging*.

38. Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory*, 20, 568–579. http://dx.doi.org/10.1080/09658211.2012.687052 doi: 10.1080/09658211.2012.687052

Additional information

Funding

This research was supported by a James S. McDonnell Foundation twenty-first century Science Initiative in Bridging Brain, Mind, and Behaviour Collaborative Award to Larry L. Jacoby and Henry L. Roediger, III. Thanks to Larry Jacoby, Henry Roediger, and Mark McDaniel for helpful comments. Thanks also to Ashley Bartels, Heather Bartels, Emily Gardner, Kim Grunde, Jennifer Kormann, Rachel Skladman, Rachel Teune, and Cecilia Votta for assistance with data collection.