Do people effectively regulate their learning of categories? In the current experiment, this question was investigated by adapting methods from Kornell and Metcalfe (2006). Two hundred and fifty-seven undergraduate students first practiced categorizing members of six artificial categories (called Fribbles). After practice, they made category learning judgments: For each category, they rated (on a scale from 0%-100%) the likelihood that they would correctly categorize new exemplars. Next, they selected half of the categories for restudy. Their selections were either honored (they restudied the selected categories) or dishonored (they restudied the unselected categories). Final test performance was greater when selections were honored, mean = .83 (SEM = .02), than dishonored, M = .73 (.02).

Participants also differed in how they used category learning judgments to select categories for restudy. Many participants selected the categories they had given lower judgments (judged as less-well-learned), but some selected the categories they had given higher judgments (judged as more-well-learned). Participants who selected the less-well-learned categories performed better on the final test when their selections were honored, M = .93 (.02), than dishonored, M = .74 (.02), but those who selected the more-well-learned categories performed better when their selections were dishonored, M = .77 (.04), than honored, M = .61 (.05). These findings suggest (1) that most people effectively regulate their learning of categories, and (2) that studying less-well-learned categories is a better strategy than studying more-well-learned categories.
LETTING STUDENTS DECIDE WHAT TO STUDY DURING CATEGORY LEARNING WILL HELP THEIR PERFORMANCE, BUT ONLY IF THEY MAKE THE RIGHT DECISIONS

A thesis submitted
To Kent State University in partial
Fulfillment of the requirements for the
Degree of Master of Arts

By

Kayla Elizabeth Morehead

August, 2017

© Copyright
All rights reserved
Thesis written by

Kayla Elizabeth Morehead

B.S., Colorado State University, 2014

M.A., Kent State University, 2017

Approved by

____ John Dunlosky ________________, Advisor

____ Maria Zaragoza ________________, Chair, Department of Psychological Sciences

____ James L. Blank ________________, Dean, College of Arts and Science
Acknowledgments

I would like to thank John Dunlosky and Nathan Foster for their help in designing and implementing this experiment. I would also like to thank Michael Mueller for his input on this experiment as well as other category learning experiments that I have conducted.
Letting Students Decide what to Study during Category Learning will help their Performance, but only if they make the Right Decisions

When learning new material such as paired associates or novel categories, people can make metacognitive judgments about how well they have learned the material and use those judgments to guide their study (for reviews, see Bjork, Dunlosky, & Kornell, 2013; Kornell & Bjork, 2007; Metcalfe, 2009). Moreover, later performance can benefit more when people guide their learning than when decisions are made for them, at least when people attempt to memorize simple items for an upcoming test of memory (Kornell & Metcalfe, 2006; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson, Dunlosky, Graf, & Narens, 1994; Rhodes, Sitzman, & Rowland, 2013; Son, 2010; Son & Metcalfe, 2000). For instance, when given a choice about which items to restudy, people generally select items that have not yet been learned (Dunlosky & Thiede, 2004). As described in further detail below, Kornell and Metcalfe (2006) demonstrated that when selecting paired associates to study for an upcoming test of memory, most participants selected items that had not been presented before instead of ones they had already studied. Critically, final performance was greater for these participants than for participants who selected items they had already studied. A major aim of the current research is to investigate whether people use metacognitive judgments to effectively regulate their learning of categories.

How can people use metacognitive judgments to regulate their learning of categories? As an example, imagine people learning to categorize birds into bird categories for a later test. They may start by making judgments for each category. For instance, they may think that they can recognize exemplars of jays, but cannot recognize exemplars of chickadees. They can then use these judgments to regulate their learning by deciding not to study any more jays, but to instead study chickadees. Is this strategy (i.e., studying the less-well-learned categories) effective, or
should they instead study the jays and stop studying the chickadees (i.e., study the more-well-learned categories)?

The effectiveness of people’s study decisions has not been investigated within the literature on category learning. Instead, when investigating category learning, the dominant approach has been to conduct experiments in which the study phase is tightly controlled by the experimenter, which has been essential for testing theoretical models of categorization (for a review, see Ashby & Maddox, 2005). However, in many contexts people do have control over their learning of categories. For example, students need to learn a variety of categories in school, including both natural (such as birds) and man-made (such as mathematics) categories. Thus, understanding how people regulate their category learning – and whether their decisions are effective – has important implications for theories of self regulation and improving students’ study behaviors.

Before describing how I evaluated whether people effectively use metacognitive judgments to guide their learning of categories, it is important to describe how monitoring of category learning is measured. To measure monitoring of category learning, researchers have participants make category learning judgments (CLJs). For each category, participants predict the likelihood that they will correctly categorize new exemplars (Jacoby, Wahlheim, & Coane, 2010). For example, people studying bird categories can make CLJs for an upcoming categorization test. They may predict that they will correctly categorize 60% of the jays, but only 20% of the chickadees. Making metacognitive judgments may help self-regulated learning because people can use these judgments to decide what to study, but the restudy decisions people make may differ depending on the type of metacognitive judgments they use to inform their decisions. Thus, it is important to contrast CLJs with other metacognitive judgments to highlight
the differences and point out why CLJs may be most relevant for self-regulated learning. To elaborate on this point, consider two other metacognitive judgments people can make while studying birds. One judgment is a judgment of learning (JOL). This judgment is made for every item. For example, people can decide that they have an 80% chance of recognizing an Oak Titmouse when it is presented on an upcoming test. They can use this judgment to decide whether they need to study that bird again, but cannot evaluate their understanding of the category structure (e.g., that an Oak Titmouse is a chickadee). Another judgment people can make is a global judgment about their overall learning of the bird categories. They may predict that they will correctly categorize 40% of all of the birds on an upcoming test. This judgment may help them decide when to stop studying overall, but it will not help them decide which birds to study. In contrast to these judgments, CLJs can help people make study decisions as the category level. Hence, CLJs are broader than JOLs, but not as broad as global judgments. Because CLJs may help students judge their knowledge of subsets of material rather than specific items or overall understanding, CLJs may be particularly relevant to education.

Prior research (Morehead, Dunlosky, & Foster, 2017) has reported that people use CLJs to regulate their learning. Given this finding, it is important to investigate whether people use CLJs to effectively regulate their learning. To investigate whether CLJs contribute to effective self regulation of category learning, I adopted methods used to evaluate whether JOLs contribute to effective study decisions when learning individual items (Kornell & Metcalfe, 2006). In second experiment (Experiment 2a) by Kornell and Metcalfe (2006), participants studied 24 word pairs. In the first phase of the experiment, only half of the word pairs were presented for study so that the other half would be unknown. Afterwards, participants selected half of the 24 word pairs for study. During restudy, participants were either shown the items they selected
(that is, their decisions were honored) or the other the items (that is, their decisions were dishonored). Final performance was greater when decisions were honored than dishonored, indicating that participants made effective study decisions.

Although Kornell and Metcalfe (2006) demonstrated that participants made effective study decisions, they also reported individual differences in self-regulation. Most participants (20 of 24) selected the unknown items for study, but the remaining four selected the items they had already studied. These four participants had the lowest final test performance and did not benefit from making their own study decisions. Performance was lower when their decisions were honored than when their decisions were dishonored. These results demonstrate the benefit of using metacognitive judgments to guide study, but also suggest individual differences in study decisions.

Individual differences in study decisions have also been demonstrated in category learning. Morehead, et al. (2017) demonstrated robust individual differences in restudy decisions when participants practiced categorizing birds. In particular, about forty percent of participants selected the less-well-learned categories for restudy, but another forty percent of participants selected the more-well-learned categories. The authors manipulated various aspects of the restudy phase in an attempt to affect how participants made selections, but individual differences in study decisions persisted across five experiments. The authors investigated why individual differences in selection occurred by comparing the two groups of participants on multiple measures (such as study performance and average CLJ). They proposed that motivation and associative memory ability may be higher for participants who selected less-well-learned categories than for those who select more-well-learned categories. Although their experiments addressed how people make selections, another important question is are both selection strategies
equally effective? Is performance similar for participants who select less-well-learned categories for restudy and participants who select more-well-learned categories for restudy?

The present experiment investigated the effectiveness of participants’ restudy decisions during category learning using the honor-dishonor method developed by Kornell and Metcalfe (2006). In order to potentially achieve the maximum honor-dishonor effect, I had to place two constraints on the method. Under ideal circumstances, the largest honor-dishonor effect is a difference of 50% on final performance between the honor and dishonor groups. To understand why that is, consider the ideal situation. Imagine two people studying categories. They finish the study phase with perfect knowledge of half of the categories (100% performance) and no knowledge of the other half (0% performance). They then select to restudy the categories for which they have no knowledge. One person is honored, studies the unknown categories, and achieves perfect knowledge on those categories (100% performance). The other person is dishonored, restudies the known categories, and does not improve knowledge of those categories. On the final test, the person who was honored scores 100% and the person who was dishonored scores 50%. Even under ideal circumstances, the largest effect possible is 50%. Hence, for the current experiment, I imposed two constraints: (1) participants continued the study phase until they had sufficient performance for half of the categories, and (2) participants selected exactly half of the categories for restudy.

During the experiment, participants practiced categorizing exemplars of artificial stimuli, called Fribbles (developed by Williams, 1997; see Figure 1). To learn each Fribble category, participants first had to figure out the rule: that each category is determined by a distinguishing feature. Once they figured out the rule, they had to determine the distinguishing feature for each category. For example, the Fribbles in the third category in Figure 1 (in the green box) all have
candy cane shaped appendages attached to their front. During study, participants were presented with pictures of individual Fribbles and categorized them with feedback. Participants continued to study Fribbles until they had correctly categorized Fibble exemplars from half of the categories. After study, participants made CLJs for all of the categories. They then selected half of the categories for restudy, and their selections were either honored or dishonored. After restudy, they were tested on all previously studied exemplars as well as new exemplars of the same categories. If people effectively regulated their category learning, then performance would be greater when participants’ selections were honored than when they were dishonored.

Given that prior research has demonstrated individual differences in study selections, I also investigated what categories participants selected for restudy and how their selections affected test performance. To investigate how different restudy selections affected performance, participants were separated into two groups: those who selected categories judged as less well learned and those who selected categories judged as more well learned. However, in the original sample (N = 80), few participants selected more-well-learned categories. In an attempt to increase the number of people in this group, I collected data from an additional 203 participants.

Individual differences in selection can lead to several outcomes. One possibility is that everyone will demonstrate an honor-dishonor effect regardless of the restudy selections that they make. This outcome would suggest that learners use strategies that are effective for them, but not necessarily effective for others. Another possibility is that some people will demonstrate an honor-dishonor effect while others will not. This outcome would suggest that only one strategy is effective for learning, but not everyone uses it. For example, selecting less-well-learned categories may be an effective strategy and demonstrate a positive honor-dishonor effect. By contrast, selecting more-well-learned categories may not be an effective strategy and
demonstrate no honor-dishonor effect or even a negative honor-dishonor effect: where dishonoring decisions improves performance. I predicted that participants who selected less-well-learned categories for restudy would show a positive honor-dishonor effect, and that participants who selected more-well-learned categories would demonstrate a negative honor-dishonor effect (similar to Kornell & Metcalfe, 2006).

Method

Participants

Two hundred and eighty-three students at Kent State University completed the experiment for course credit in introductory psychology or research methods. Twenty-six participants did not complete the experiment. So the final sample consisted of 257 participants. One hundred and thirty participants were randomly assigned to the honor condition, and 127 to the dishonor condition.

Materials

The stimuli used were artificial categories called Fribbles. Fribbles are three-dimensional computerized objects that are meant to represent natural categories, but were novel to participants. Each Fribble has a body with four appendages attached to it that can vary (see Figure 1). Each category is defined by a single appendage that is the same for all exemplars. To learn categories (six categories total: Beme, Coge, Kipe, Tyfe, Vune, and Zade), participants needed to discover and remember which appendage distinguished one category from the others. Category names were randomly assigned to Fribble categories for each participant.

Procedure

Participants completed the study on a computer in a laboratory setting in one, 30-minute session. Some participants could not complete the study within 30 minutes and either took
longer, upwards of 60 minutes, or did not finish the study. Participants first completed an active, self-paced study phase. Six Fribbles per category were presented (36 Fribbles total) in a random order with the constraint that three Fribbles from the same category could not be presented in a row. On each screen during the study phase, participants were presented with a Fribble, all six category names, the prompt: “What kind of Fribble is this?”, and a box to type their answer. They typed the name of the category to which they thought the Fribble belonged. Participants were then presented with the correct name of the category and clicked a button at the bottom of the screen to view the next Fribble. Participants continued to study until they reached criterion: correctly categorized three exemplars of a category in a row for three of the categories. If a participant did not reach criterion after all Fribbles had been presented once, they repeated the trial with the same Fribbles presented in the same order.

After the study phase, participants made CLJs for all six categories. Specifically, they were told, “You will see the name of a Fribble category and be asked to judge (on a scale from 0%-100%) how well you believe you could categorize novel Fribbles into their correct category.” Each category name was presented individually and participants entered a value between 0 and 100. Then they selected exactly three of the categories for restudy. Category names were shown individually in alphabetical order, and participants selected “yes” to restudy a category or “no” to not. A counter at the bottom of the screen showed participants how many categories they had left to select. If participants did not select three categories when all names had been presented once, they were told to select more categories, and the restudy-selection phase started over with the first name of those categories that had not been selected. When participants selected three categories, they began the restudy phase. During this phase, the honor group studied Fribbles from the three categories they had selected, and the dishonor group
studied Fribbles from the three categories they had not selected. The restudy phase was similar to the study phase except that the Fribbles were presented in a newly randomized order, only half of the categories were presented, and all of the studied Fribbles (six Fribbles per category, eighteen total) were presented three times. Immediately after the restudy phase, participants completed the final test. The test was similar to the study phase except that no feedback was given, and the stimuli were presented in a newly randomized order. Participants were tested on the same 36 Fribbles that had already been presented as well as 36 new Fribbles (six per category).

Results

Honor-Dishonor Effect

To investigate the honor-dishonor effect, I compared final test performance for the honor and dishonor groups. The results are presented in Figure 2, and means and standard errors of the mean are presented at the top of Table 1. Overall final test performance was greater for the honor group than the dishonor group, $t(255) = 3.57, p < .001, d = .45$. Final test performance for novel categories was also greater for the honor group than the dishonor group, $t(255) = 3.24, p < .01, d = .41$, as was final test performance for studied categories, $t(255) = 3.78, p < .001, d = .47$. Overall, the results demonstrate a positive honor-dishonor effect for novel and studied categories.

CLJ by Selection Correlation

To investigate which categories participants selected for restudy, intra-individual gamma correlations (Nelson, 1984) between CLJs and restudy selections were computed. For each participant, we compared their CLJ for each Fribble category and whether that category was
selected for restudy\(^1\). Values of this correlation can range from -1.0 to 1.0. In this case, a negative value (closer to -1.0) would indicate that a participant selected categories for restudy that had been judged as less-well-learned (more difficult) compared to the other categories. By contrast, a positive value (closer to 1.0) would indicate that a participant selected categories for restudy that had been judged as more-well-learned (easier) compared to the other categories.

The mean correlation between CLJs and study selections across participants was -0.42 \((SEM = .05)\), suggesting that many participants selected the less-well-learned categories for restudy. However, the standard deviation of the correlations was 0.71, and correlations ranged from -1.0 to 1.0, suggesting that participants differed widely in how they made study selections, which I address next.

**Individual Differences in Category Selection**

To investigate individual differences in restudy selections, a frequency distribution of participants’ correlations between CLJs and selections was constructed. To build the frequency distribution, correlations were binned in the following way: correlations of -1.0 were in the first bin, correlations between -0.99 and -0.8 (inclusive) were in the next bin, correlations between -0.79 and -0.6 were in the third bin (and so on) until the final bin, which included correlations of 1.0. As evident from Figure 3, and in accordance with the mean correlation, the majority of participants had negative correlations between their CLJs and selections. However, some participants had positive correlations, indicating that they selected more-well-learned categories rather than the less-well-learned ones. Are these two strategies (selecting less-well-learned categories and selecting more-well-learned categories) equally effective, or is one more effective than the other?

\(^1\) Gamma correlations could not be computed for thirty participants because these participants gave the same CLJ to every category.
Honor-dishonor effect as a function of selection strategy.

To compare the effectiveness of the two strategies, I conducted the honor-dishonor analysis for those who selected less-well-learned categories and those who selected more-well-learned categories. Before conducting the analysis, participants were grouped into the two selection groups. To create these groups, I used the same approach as Morehead, et al. (2017). Anyone with a correlation between CLJs and selections at or below -0.5 was classified as selecting less-well-learned categories (honor: 71 participants; dishonor: 72 participants), and anyone with a correlation at or above 0.5 was classified as selecting more-well-learned categories (honor: 18 participants; dishonor: 22 participants). Anyone with a correlation between -0.5 and 0.5 was excluded from further analyses because I was not able to discern whether participants in this group selected categories using another strategy or selected categories at random. Forty-two participants were in this group.\(^2\) The conservative cutoffs of -0.5 and 0.5 were chosen to increase the likelihood that participants within these groups applied one of the two selection strategies. By contrast, values closer to zero could indicate that participants used one of those strategies, used another strategy, or randomly selected categories (e.g., someone with a correlation of 0.2 could have selected more-well-learned categories or could have used another strategy where some categories selected happened to be more-well-learned).

Three 2 (selection: less-well-learned, more-well-learned) X 2 (group: honor, dishonor) ANOVAs were conducted for overall final test performance, final test performance for novel exemplars, and final test performance for studied exemplars. Results are presented in Figure 4, and means and standard errors of the mean are presented in Table 1.

\(^2\) I also conducted these analyses splitting the two groups at 0, thus including all participants. Results were similar to those reported here.
A significant main effect of selection occurred on overall final test performance, 
\[ F(1, 179) = 16.11, p < .001, \text{partial } \eta^2 = .083, \] such that final test performance was greater for those who selected less-well-learned categories than for those who selected more-well learned categories. Also, a significant main effect of selection behavior occurred for novel exemplars, 
\[ F(1, 179) = 15.65, p < .001, \text{partial } \eta^2 = .080, \] and for studied exemplars, \[ F(1, 179) = 14.93, p < .001, \text{partial } \eta^2 = .077. \] A non-significant main effect of honor-dishonor group occurred on overall final test performance, for novel exemplars, and for studied exemplars \((F < 1)\).

Most important, a significant interaction between study selection group and honor-dishonor group occurred on overall final test performance, \[ F(1, 179) = 26.55, p < .001, \text{partial } \eta^2 = .129. \] The interaction was also significant for novel exemplars, \[ F(1, 179) = 24.27, p < .001, \text{partial } \eta^2 = .119, \] and for studied exemplars, \[ F(1, 179) = 26.38, p < .001, \text{partial } \eta^2 = .128. \] As evident from Figure 4, for those who selected less-well-learned categories, final test performance was significantly higher when their restudy selections were honored than when they were dishonored. By contrast, for those who selected more-well-learned categories, final test performance was significantly lower when their restudy selections were honored than when they were dishonored.

**Differences between groups.**

A secondary question regarding individual differences in selection is why do some participants select less-well-learned categories and others select more-well-learned ones? Answering this question could explain why selecting less-well-learned categories was a more effective strategy in this context. For example, participants who selected less-well-learned categories may have been more motivated to do well on the task.

To investigate how these two groups may differ, those who selected
less-well-learned categories were compared to those who selected more-well-learned categories on a variety of measures. Results are presented in Table 2 and discussed below.

One variable participants may differ on is study performance. That is, participants who performed poorly during the initial study phase may have selected more-well-learned categories for restudy in an attempt to improve performance for at least some easier categories. By contrast, those who performed well during the study phase may have selected the less-well-learned categories to further improve performance. However, study performance did not significantly differ between the two groups ($p = .22$; see row Study Performance in the table), but note that based on the design, participants had to reach criterion of about 50% performance during the study phase (as shown in Table 2, performance for both groups was close to 50%). Hence, it is not surprising that groups did not differ on study performance, but they may differ on the number of study trials they completed before reaching criterion (where one trial is once through all of the studied Fribbles). Although the difference was not significant ($p = .08$; row Number of Study Trials), about one more trial was used to complete the study phase by participants who selected more-well-learned categories than participants who less-well-learned categories.

The groups also differed on both restudy performance ($p < .001$; row Restudy Performance) and final test performance ($p < .01$; row Test Performance), such that performance was greater for those who selected less-well-learned categories than those who selected more-well-learned categories. Because conditions were manipulated during the restudy phase, these results are difficult to interpret, but they do suggest that those who selected less-well-learned categories may have learned more than those who selected more-well-learned categories despite similar performance during the study phase.
Finally, the groups were compared on average CLJs across categories and gamma correlation between study performance and CLJs. Although performance during the study phase was similar for both groups, perceived performance may have been lower for those who selected more-well-learned categories than for those who selected less-well-learned categories. This would be evident from differences in average CLJs. If participants who selected more-well-learned categories were less confident, they would give overall lower CLJs. As evident from row *Average CLJ* in Table 2, average CLJs were significantly lower for participants who selected more-well-learned categories than for participants who selected less-well-learned categories (*p* < .05). The gamma correlation between study performance and CLJs measures the accuracy of CLJs in assessing study performance. The two groups may differ on these values if one group was inaccurate at judging performance. In this case, using CLJs to make restudy decisions would most likely not benefit performance. The groups did not significantly differ on this measure (*p* = .21; row *Study Performance by CLJ*). The results of these exploratory analyses suggest that participants who perform better on the task are more likely to select less-well-learned categories for restudy than more-well-learned categories.

**General Discussion**

The present experiment was the first to investigate whether people effectively use metacognitive judgments when making restudy selections during category learning. Participants studied artificial categories (called Fribbles), judged their learning of each category, and made restudy selections. To investigate the effectiveness of those selections, participants’ selections were either honored (they restudied the selected categories) or dishonored (they restudied the unselected categories). Overall, participants demonstrated a positive honor-dishonor effect; Final test performance was greater when participants’ selections were honored (*mean* = .83) than
when they were dishonored (mean = .73). The results suggest that most participants make effective study selections while learning categories.

Although many participants made effective study selections, individual differences in selection behavior also occurred. Specifically, many participants selected categories they had judged as less well learned for restudy, but some participants selected categories that they had judged as more well learned. Participants who selected less-well-learned categories for restudy demonstrated a positive honor-dishonor effect (final test performance was greater when selections were honored than dishonored), but participants who selected more-well-learned categories demonstrated a negative honor-dishonor effect (final test performance was greater when selections were dishonored than honored). These results suggest that CLJs can support effective study decisions during category learning, but only if people make appropriate control decisions.

**Why do Individual Differences Occur?**

To evaluate this question, the selection groups were compared on multiple measures as in Morehead, et al. (2017). In their analysis, Morehead, et al. (2017) reported few differences between the groups, except for small differences in associative memory ability and the number of categories selected for study, favoring participants who selected less-well-learned categories. In the current experiment, the selection groups differed on the number of trials to reach study phase criterion, such that fewer trials were needed to reach criterion for participants who selected less-well learned categories (than for participants who selected more-well-learned categories). Also, performance during the restudy phase and on the final test was greater and average CLJs were higher for participants who selected less-well-learned categories. These results suggest that the task was easier for participants who selected less-well-learned categories, and they perceived
that they did better on the task. Because they performed well on the task, they may have felt that they could select the less-well-learned categories to further improve their performance. By contrast, the task was more difficult for participants who selected more-well-learned categories, and they perceived that they did worse on the task. Because they did not perform as well on the task, they may have selected more-well-learned categories in an attempt to improve performance for some of the easier categories. Although selecting more-well-learned categories may have been a rational decision, it did not benefit them. Performance was still better when they were dishonored than when they were honored.

Another potential explanation for why differences in selection occurred is that the two groups used different agendas. One agenda, based on the discrepancy-reduction model (Dunlosky & Thiede, 1998), predicts that people will allocate more effort to items furthest from their goal state (the most difficult items), particularly when they have a mastery goal. This agenda may have been adopted by participants who selected less-well-learned categories in the current experiment. These participants may have had a mastery goal for the final test, and hence selected the most difficult categories in an attempt to master all of the categories. The other agenda, based on the region of proximal learning (Metcalfe & Kornell, 2005), proposes that people will select the easiest of the unlearned items for restudy. This agenda may have been adopted by participants who selected more-well-learned categories in the current experiment. Although this agenda was not effective in the current experiment, it may be effective when categories are hierarchically related, such that the easiest categories need to be learned before learning more difficult categories. For example, math concepts are often related hierarchically (e.g., students must learn how to solve linear equations before they can learn how to solve quadratic ones). Hence, a student with low knowledge may benefit more from studying linear
equations before studying quadratic equations. However, Fribble categories are not related hierarchically. A person does not necessarily need to learn easier Fribble categories before learning more difficult ones. Hence, studying more-well-learned categories is not an effective agenda when studying Fribble categories, but it may be effective in a situation where categories are hierarchically related.

**Do these Effects Extend to Other Material?**

I just completed another experiment conceptually replicating the current one with different stimuli. In that experiment, participants practiced categorizing pictures of birds into bird categories. The bird stimuli differ from the Fribbles in that they are natural categories and cannot be categorized based on a simple rule. I chose these stimuli because past research (Morehead, et al., 2017) demonstrated robust individual differences in restudy decisions; About forty percent of participants selected less-well-learned categories, and another forty percent selected more-well-learned categories. I anticipated that using these materials would increase the sample size of participants who selected more-well-learned categories. However, unlike results from Morehead et al. (2017), more participants selected less-well-learned categories (n = 42) than more-well-learned categories (n =23).

Preliminary results from this experiment replicated the results of the present experiment with Fribbles, but the effects were small. Final test performance did not significantly differ between the honor group, $M = .68$ ($SEM = .02$), and the dishonor group, $M = .67 (.01)$, $t (78) = 0.31, p = .76$. Within the subset of participants who selected less-well-learned categories, final test performance was non-significantly greater for the honor group,
\( M = .71 (.02) \), than the dishonor group, \( M = .67 (.02) \), \( t (40) = 1.38, p = .18 \). Within the subset of participants who selected more-well-learned categories, final test performance did not significantly differ between the honor group, \( M = .66 (SEM = .03) \), and the dishonor group, \( M = .66 (.04), t (21) = 0.95, p = .96 \). One possible reason why the effects were smaller in this experiment is that the conditions necessary to obtain a robust honor-dishonor effect were not met. For instance, one issue may be that mean performance on the study phase was 59%. As stated in the introduction, performance on the study phase should be around 50% to potentially maximize the honor-dishonor effect. One way to address this issue would be to lower performance to around 50% by using three normatively easy bird categories and three normatively difficult ones. To evaluate this possibility as well as others, I will analyze the data to investigate where the method was not sensitive enough to maximize the honor-dishonor effect and design a new experiment based on the results of that investigation.

**Conclusion**

Overall, the results of the current experiment suggest that most people make effective study selections during category learning when they use monitoring judgments to make study selections. However, not everyone makes effective study selections. People who do not make effective study selections may do so in an attempt to compensate for low performance, but this strategy is not effective. Instead, selecting less-well-learned categories is a better study strategy for all people. Although research with educationally relevant material is needed, studying less-well-learned material may also be a more effective strategy for studying material for courses.
References


Table 1. *Final test performance as a function of selection behavior and honor-dishonor group.*

<table>
<thead>
<tr>
<th></th>
<th>Honor</th>
<th>Dishonor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>.83 (.02)</td>
<td>.73 (.02)</td>
</tr>
<tr>
<td>Novel</td>
<td>.82 (.02)</td>
<td>.72 (.02)</td>
</tr>
<tr>
<td>Studied</td>
<td>.84 (.02)</td>
<td>.74 (.02)</td>
</tr>
<tr>
<td><strong>Less-Well-Learned</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>.93 (.02)</td>
<td>.74 (.03)</td>
</tr>
<tr>
<td>Novel</td>
<td>.92 (.02)</td>
<td>.72 (.03)</td>
</tr>
<tr>
<td>Studied</td>
<td>.94 (.02)</td>
<td>.74 (.02)</td>
</tr>
<tr>
<td><strong>More-Well-Learned</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>.61 (.05)</td>
<td>.77 (.04)</td>
</tr>
<tr>
<td>Novel</td>
<td>.58 (.05)</td>
<td>.76 (.05)</td>
</tr>
<tr>
<td>Studied</td>
<td>.64 (.06)</td>
<td>.79 (.04)</td>
</tr>
</tbody>
</table>

*Note. Standard errors of the mean are in parentheses.*
Table 2. Comparison between groups on multiple measures.

<table>
<thead>
<tr>
<th></th>
<th>Less-Well-Learned</th>
<th>More-Well-Learned</th>
<th>Significance Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Performance</td>
<td>.48 (.01)</td>
<td>.45 (.02)</td>
<td>( p = .22 )</td>
</tr>
<tr>
<td>Number of Study Trials</td>
<td>2.08 (.11)</td>
<td>3.03 (.51)</td>
<td>( p = .08^a )</td>
</tr>
<tr>
<td>Restudy Performance</td>
<td>.91 (.01)</td>
<td>.81 (.02)</td>
<td>( p &lt; .001^a )</td>
</tr>
<tr>
<td>Test Performance</td>
<td>.84 (.02)</td>
<td>.72 (.04)</td>
<td>( p &lt; .01 )</td>
</tr>
<tr>
<td>Average CLJ</td>
<td>60.61 (1.59)</td>
<td>52.87 (3.46)</td>
<td>( p &lt; .05 )</td>
</tr>
<tr>
<td>Study Performance by CLJb</td>
<td>.53 (.03)</td>
<td>.44 (.07)</td>
<td>( p = .21 )</td>
</tr>
</tbody>
</table>

Note. Standard errors of the mean are in parentheses. \(^a\)Levene’s test of equal variances was significant, so these t-tests were conducted not assuming equal variances. \(^b\)This measure is the gamma correlation between study performance and CLJs.
Figure 1. From Carvalho and Goldstone (2014). Exemplars of members of each of the six categories. Each colored box represents Fribbles from one category.
Figure 2. Final test performance for honor and dishonor groups.
Figure 3. Frequency distribution of within-participant correlations between CLJs and restudy selections.
Figure 4. Final test performance for honor and dishonor groups as a function of whether participants selected categories for restudy that were judged as less-well-learned or more-well-learned.