Do people use category-learning judgments to regulate their learning of natural categories?

Kayla Morehead¹ · John Dunlosky¹ · Nathaniel L. Foster²

Published online: 13 July 2017
© Psychonomic Society, Inc. 2017

Abstract Although research has established that people can accurately judge how well they have learned categories, no research has examined whether people use their category-learning judgments (CLJs) to regulate their restudy of natural categories. Thus, in five experiments we investigated the relationship between people's CLJs and selections of categories for restudy. Participants first attempted to learn natural categories (bird families; e.g., finches, grosbeaks, and warblers) so that they could categorize new exemplars on a final test. After this initial study phase, participants made a CLJ for each category and then selected a subset of the categories for restudy. Across experiments, we also manipulated several variables (e.g., selecting either three or nine categories, or obtaining 30% vs. 80% performance on the final test) that were expected to influence restudy selections. However, the manipulations typically had minimal impact. More important, in all experiments we found an unexpected outcome: Some participants tended to select the categories they judged to be most well learned for restudy, and others tended to select those judged to be least well learned. We discovered these qualitative differences in the use of CLJs to make restudy selections by using post-hoc analyses in Experiments 1a and 1b, and hence we sought to (a) replicate them in Experiments 2, 3, and 4 and (b) provide preliminary evidence regarding factors that can (vs. cannot) account for them. Most important, evidence across all of the experiments supported the conclusion that people do use their CLJs to select categories for restudy.

Keywords Selfregulation · Metacognition · Category-learning judgments · Individual differences

How do people decide what to study when they are learning categorical information? For instance, when learning to classify diseases, do medical doctors choose to study diseases that they are having the most difficulty learning to diagnose? When learning to classify plants during job training, do park rangers choose to study the plants they are having the most difficulties classifying? And, while learning various categories in subjects such as geology, biology, or math, do students consider how well they have already learned the categories when deciding which ones to restudy? As a specific example relevant to the present research, imagine people learning to identify different bird families—such as grosbeaks, finches, and sparrows—so they can later identify those birds in the field. As they study, they can monitor their learning of the different bird families by judging how well they can identify different species within a family. In this case, do they use their category-learning judgments to decide which bird families to restudy? And, if so, what types of study decisions do they make on the basis of their judgments? Answering these questions is our central aim.

How well people monitor their category learning has recently received attention in the field (DeSoto & Votta, 2016; Doyle & Hourihan, 2016; Hartwig & Dunlosky, 2017; Jacoby, Wahlheim, & Coane, 2010; Tauber & Dunlosky, 2015; Thomas, Finn, & Jacoby, 2016; Wahlheim & DeSoto, 2017; Wahlheim, Dunlosky, & Jacoby, 2011; Wahlheim, Finn, & Jacoby, 2012; Yan, Bjork, & Bjork, 2016; Zauhar, Bajšanski, & Domijan, 2016). This research has focused on estimating the accuracy of people's category-learning judgments and is based on the assumption that monitoring category learning plays a functional role in self-regulated learning.

✉ John Dunlosky
jdunlosk@kent.edu

¹ Kent State University, Kent, OH, USA

² St. Mary's College of Maryland, St. Mary's City, MD, USA

For instance, Wahlheim, Finn, and Jacoby conjectured that “accurate assessments of differences in the extent to which people have learned various topics (e.g., attention, memory, problem solving) might serve to guide the allocation of additional study” (p. 704). Likewise, Thomas, Finn, and Jacoby assumed that people “judge their knowledge of a given topic and then use that judgment to choose which topics to emphasize during study” (p. 2). Despite the importance of this assumption (i.e., that people use their judgments in making restudy decisions) to justifying improving the accuracy of people’s category monitoring, no research has evaluated whether people use their monitoring of category learning to regulate subsequent study. Thus, in the present experiments we explored whether a relationship exists between peoples’ judgments of category learning and their self-regulated learning of natural categories.

To lay out our approach, we first need to describe how monitoring of category learning is measured, because we adapted the methods used in prior monitoring research to investigate whether people use this monitoring to make restudy decisions. Participants typically first study categories and then make *category-learning judgments* (CLJs), which are metacognitive judgments made at the category or concept level.¹ CLJs were introduced by Jacoby, Wahlheim, and Coane (2010), who had participants learn to categorize pictures of various bird species (exemplars) into 12 bird families (categories). After studying exemplars of all 12 categories, participants were shown each of the category names individually and asked to predict future categorization performance for novel exemplars of each category. Then, they completed a final test over studied and novel exemplars. Each participant’s CLJs and test performance were correlated across categories, and the mean correlation was above chance (greater than zero) but far from perfect, which indicates that participants have some ability to accurately assess their learning of natural categories. Other researchers have also reported above-chance levels of CLJ accuracy (see aforementioned citations), but none have evaluated whether people use their CLJs when selecting categories for restudy. Most important, to address this main issue, we used a standard method to obtain CLJs (as per Jacoby et al., 2010) with one key addition. In particular, after participants made CLJs, they were instructed to select a subset of the 12 categories for restudy.

Will people use CLJs to make their restudy selections? According to the agenda-based regulation framework (Ariel, Dunlosky, & Bailey, 2009; Dunlosky & Ariel, 2011; see also Nelson & Narens, 1990; Winne & Hadwin, 1998), people develop and use agendas—or simple plans—to decide which categories to select for restudy. In particular, people presumably form an agenda that they believe will help them reach their learning goal most efficiently and hence would be expected to use their CLJs to make restudy decisions. For instance, in contexts in which learners adopt a mastery goal and are given ample time to study the to-be-learned items, they tend to study the most difficult, unlearned items. This agenda (study the majority of difficult, unlearned items) is most consistent with the discrepancy reduction model (Dunlosky & Thiede, 1998), which predicts that people will allocate more effort to items furthest from their goal state (or the most difficult items). However, if learners are not given enough time to study, or if their goal emphasizes efficiency over mastery, they tend to develop an agenda to select the easier unlearned items for restudy (Son & Metcalfe, 2000; Thiede & Dunlosky, 1999), consistent with the region-of-proximal-learning model (Metcalfe & Kornell, 2005). This agenda emphasizes efficiency in a situation in which learners cannot master all of the content. These expectations have been evaluated in the memory literature, in which learners select and restudy items for an upcoming memory test. In particular, when learners have a mastery goal for learning simple associates, they tend to select the more difficult pairs for restudy, but when they are told their goal is to learn just a few of the pairs (e.g., six out of a list of 30), they shift to selecting only a few of the easier pairs for restudy (for details, see Dunlosky & Thiede, 2004). Given the evidence from the memory literature that people make different decisions depending on the task constraints, we manipulated some constraints (described below) to evaluate whether these would also influence how people make restudy decisions when learning categories.

Experiments 1a and 1b

For Experiments 1a and 1b, we had participants practice categorizing pictures of birds into 12 categories. After study, participants made a CLJ for each category and then selected a subset of categories for restudy. As one task constraint, we varied the number of categories participants were allowed to select. One group of participants was instructed to select three of the 12 categories for restudy, and the other was instructed to select nine. Consistent with prior research (e.g., Dunlosky & Thiede, 2004), we predicted that if participants could select only three categories, they would tend to set a lower performance goal and hence select the easier categories (i.e., those given higher CLJs) in an attempt to increase the chances that they would learn at least some of the categories. By contrast, if

¹ Unlike item-by-item judgments, which are judgments about the ability to memorize individual items, CLJs are judgments about the ability to generalize to new category exemplars. For the latter type, learners are asked to judge how well they can extrapolate the features of category exemplars (rather than judge how well they have memorized exemplars). CLJs also differ from global judgments, which involve judging performance for the overall task (and not just judging subsets of the task—or categories—as with CLJs). Thus, the different types of judgments are functionally distinct, but the degree to which these judgments are psychologically distinct is an empirical question that is not addressed here.

participants could select nine, they would tend to set a higher performance goal and hence select the more difficult categories (i.e., those given lower CLJs). These predictions are based on evidence from memory research (e.g., deciding which paired associates to restudy), and hence they may not generalize to how people regulate their learning of categories.² One possibility is that people might use a different strategy (rather than using CLJs) to make restudy selections. For instance, they may select categories that they want to compare to each other so as to identify differences between them. If so, regardless of the task constraints, people might not use their CLJs to decide which categories to restudy.

In both experiments, we also manipulated a second variable during restudy selection. For Experiment 1a, during the selection phase, CLJs were presented above each corresponding category name for some participants and were absent for the other participants. Our rationale for this manipulation was that if people use their CLJs spontaneously, no differences will occur between the CLJ present and absent groups. However, if people do not use their CLJs spontaneously, then presenting them during selection could increase the likelihood that people use them to make selections. If so, a relationship between CLJs and selections would occur in the CLJ-present group (positive or negative, based on whether participants were in the select-three or select-nine group) but not in the CLJ-absent group. For Experiment 1b, during the selection phase we manipulated how the categories were presented for selection. Specifically, in one condition all the category names were presented simultaneously in an array (as in Exp. 1a), and in the other they were presented sequentially for selection. Our rationale for this manipulation was based on work by Dunlosky and Thiede (2004), who found that presenting items simultaneously for selection increased the likelihood that participants would develop and implement an effective plan for selecting items for restudy. If the same held in the present case, then the relationship between CLJs and restudy selections would be stronger when categories were presented simultaneously rather than sequentially.

To foreshadow, the results from Experiment 1a were as expected. That is, those who could select three categories for restudy demonstrated a positive relationship between CLJs and selections, and those who could select nine demonstrated a negative relationship. However, the effects were small, so we conducted a replication in Experiment 1b. We expected that the task constraints would strongly influence how

participants made restudy selections, on the basis of prior evidence from the memory literature using verbal materials (for reviews, see Dunlosky & Ariel, 2011; Son & Metcalfe, 2000), but all of the manipulations tended to have a minor—if any—impact on how people used CLJs during restudy selection (hence, we report the relevant outcomes below only briefly). Instead, a novel outcome was that relatively extreme individual differences appeared in the use of CLJs during selection—an outcome that we investigated in Experiments 2, 3, and 4.

Method

Participants A total of 129 undergraduates at Kent State University completed Experiment 1a for course credit in introductory psychology or research methods. Fourteen of these participants were removed due to a programming error, leaving 115 participants in the final data set. A further 123 Kent State University undergraduates completed Experiment 1b. Nine of these participants were removed due to a programming error, leaving 114 participants.

Materials Our stimuli were images of birds used by Wahlheim, Finn, and Jacoby (2012). The birds were presented individually on a brown background, and the images were resized so all of the birds were about the same size (for examples, see Wahlheim et al., 2012). Each individual bird represented a bird species within a bird family. Participants studied 12 bird families (chickadee, finch, flycatcher, grosbeak, jay, oriole, sparrow, swallow, thrasher, thrush, vireo, and warbler). We used 12 images per bird family. Six birds were presented at study and test, and six others were presented only at test.

Procedure During the study phase, participants were shown an individual bird and chose the category they thought it belonged to from a list of all 12 category names. Participants then were shown the correct name (along with the image) and clicked a button to view the next bird. Birds were presented in a random order, with the constraint that no more than two birds from the same category could be presented in a row. This study phase consisted of three blocks, with each block containing all 72 birds.

After the study phase, participants made a CLJ for each category. The category names were presented individually in alphabetical order with the prompt “What is the likelihood that you will correctly classify NOVEL birds from this family?” Participants made CLJs on a scale from 8% (*guessing*) to 100% by typing the number into a box and then pressing the enter key.

After participants had made the CLJs, they selected a subset of categories for restudy. The category names were

² Given that selection occurs at the categorical level, other predictions could be developed, depending on the level of performance achieved for any given category. For instance, even when participants selected nine categories, they might still select the easier categories if the level of performance for all categories was relatively low. We do not provide details for all of these possibilities, because (to foreshadow) the focal manipulations that we consider here had a limited impact on how people used CLJs to select categories, and hence none of the predictions were consistently confirmed.

presented simultaneously in alphabetical order. In Experiment 1a, participants' CLJs were presented over each category name for one group ($n = 60$), and CLJs were not presented for the other ($n = 56$). In Experiment 1b, the category names were presented simultaneously for one group ($n = 63$) and sequentially for the other ($n = 51$). The simultaneous display was the same as in Experiment 1a. For the sequential group, the names were presented one at a time in alphabetical order. Participants indicated whether or not they wanted to restudy a category by clicking a “yes” or a “no” button at the bottom of the screen. For both experiments, a counter at the bottom of the screen told participants how many families they had left to select. In all, 56 of the participants in Experiment 1a and 60 participants in Experiment 1b could select exactly three categories, and the remainder (Exp. 1a, $n = 60$; Exp. 1b, $n = 54$) could select exactly nine. After the selection phase the participants did not restudy, but instead proceeded directly to the final test. We excluded the restudy phase in order to replicate previous work that had estimated CLJ accuracy by comparing CLJs to test performance. For the test, participants classified the same 72 birds from the study phase in a newly randomized order, followed by 72 novel birds.

Results and discussion

Selection as a function of category-learning judgments To evaluate whether participants were biased toward selecting categories they judged to be more (or less) well learned, we computed a within-participant gamma correlation between each participant's CLJs and restudy selections. Gamma is a nonparametric correlation (for details, see Nelson, 1984) that indicates the degree to which CLJs are monotonically related to restudy selections (so the absolute level of the correlation could be near 1.0 even if CLJs and restudy selections were not linearly related). Negative correlations would indicate that the participants tended to select categories they judged to be more difficult; positive correlations would indicate that they tended to select categories they judged to be easier; and near-zero correlations would suggest that they did not use CLJs to select categories for restudy. (An alternative approach to analyses would be to compute the mean CLJs for the categories that were vs. were not selected; as is shown in the Appendix, this approach yielded outcomes that were consistent with the correlational analyses described in the text.) The means across participants' correlations for Experiments 1a and 1b are presented in Table 1.

First, consider the outcomes from Experiment 1a (top half of Table 1): Those who selected nine categories tended to select the categories that they had judged to be more difficult, whereas those who selected three categories tended to do the opposite. Consistent with this observation, a 2 (number selected: 3

Table 1 Correlations between category-learning judgments (CLJs) and category selections

Group	Select Three	Select Nine	Group Means
Experiment 1a			
CLJ present	.30 (.14)	-.14 (.15)	.08 (.11)
CLJ absent	.10 (.18)	-.24 (.15)	-.08 (.12)
Group means	.21 (.11)	-.19 (.11)	.01 (.08)
Experiment 1b			
Simultaneous	-.10 (.14)	.12 (.16)	.00 (.10)
Sequential	-.23 (.15)	.01 (.16)	-.11 (.11)
Group means	-.15 (.10)	.07 (.11)	-.05 (.08)

Main entries represent means across within-participant correlations between CLJs and category selections. Positive values indicate a bias to select the categories judged as easier, and negative values indicate a bias toward selecting categories judged as being more difficult. Entries in parentheses are the corresponding standard errors of the means.

vs. 9) $\times$ 2 (CLJ presence: present vs. absent) full factorial analysis of variance (ANOVA) revealed a main effect of number of categories selected, $F(1, 113) = 6.34$, $MSE = 0.68$, $p < .05$, $\eta_p^2 = .06$. The main effect of CLJ presence and the interaction were not significant, $F_s < 1.0$. Note, however, that our main question was not whether the correlations from the select-three and select-nine groups would differ from each other, but instead whether they would differ from zero, which would suggest that people used the CLJs to make their restudy selections. The correlations between CLJs and restudy selections were not significantly different from zero, but they trended in the predicted directions for the select-three group, $t(54) = 1.92$, $p = .06$, and for the select-nine group, $t(58) = 1.76$, $p = .08$.

Because the main effect was small, we attempted a replication and extension in Experiment 1b. Those values are reported in the bottom half of Table 1. The trend apparent in Experiment 1a was not evident in Experiment 1b, and the main effect of number selected from the 2 (number selected) $\times$ 2 (selection format: simultaneous vs. sequential) ANOVA was not statistically significant, $F(1, 107) = 2.20$, $MSE = .65$, $p = .14$, $\eta_p^2 = .02$. The main effect of selection format and the interaction were also not significant, $F_s < 1.40$. Moreover, as in Experiment 1a, the correlations for the select-three and select-nine groups were not significantly different from zero for the select-three group, $t(57) = 1.53$, $p = .13$, or for the select-nine group, $t(52) = 0.50$, $p = .55$.

Because the mean correlations within groups were small, we also computed the mean correlations between CLJs and selections across groups: Those values are presented in the last column of Table 1. Similar to the within-group data, the overall mean correlations suggest that many participants may not have used the CLJs when deciding which categories to restudy.

Fig. 1 Frequency distribution of correlations between category-learning judgments (CLJs) and selections for Experiment 1a

Frequency distributions Although the mean correlations indicate a minor relationship between CLJs and restudy selections, strong relationships could be evident when one considers individual differences in selection behavior. To investigate this possibility, we constructed frequency distributions of the participants' correlations between CLJs and selections. We binned correlations for the frequency distributions in the following way: Scores of -1 were grouped in the first bin, scores between $-.99$ and $-.9$ (inclusive) were in the next bin, scores between $-.89$ and $-.8$ were in the third bin, and so on, until the final bin, which included correlations of 1 . Given that the mean correlations were near zero, one expectation was that the frequency distributions would be normally distributed

and centered near zero. However, as is evident from inspecting Figs. 1 and 2, the correlations in Experiments 1a and 1b were bimodally distributed (note that even for Exp. 1a, in which the select-three and select-nine groups differed significantly with respect to their mean correlations, the correlations within both groups demonstrated the extreme bimodal distribution; for brevity, we do not present distributions as a function of the secondary variables, but these conditional analyses are available from the first author). These outcomes—with individual differences in selection strategies, in which some participants selected categories they had judged to be easier and others selected categories that they had judged to be more difficult—were unexpected and required replication.

Fig. 2 Frequency distribution of correlations between category-learning judgments (CLJs) and selections for Experiment 1b

Selection group differences Because the individual differences in restudy selection were unexpected, in this section we consider individual differences within Experiments 1a and 1b that could possibly explain why some participants selected easier categories for restudy, whereas others selected more difficult categories. To perform this analysis, correlations between CLJs and selections greater than or equal to .50 formed the *easy group*, whereas correlations less than or equal to $-.50$ formed the *difficult group* (see Table 2 for the sample size of each group). (For brevity, we use the terms “easy” and “difficult” to refer to “selected categories judged to be more well learned” and “selected categories judged to be less well learned,” respectively.) We chose these cutoffs to increase the chance that the participants in each group would likely be those who applied one of the strategies (e.g., selected either easy or difficult categories). In particular, a CLJ–selection correlation of 1.0 would indicate that a participant exclusively selected the easy over the difficult categories, and a CLJ–selection correlation of -1.0 would indicate that a participant exclusively selected the difficult over the easy categories (i.e., any divergence would yield a less extreme correlation; see Nelson, 1984, for the computational details). Correlations closer to zero (and even correlations of 0) could indicate that a participant used one of these strategies (e.g., a $-.20$ could represent selecting difficult categories), but they also could mean that the participant used a mixture of the strategies or relied on another strategy. That is, correlational values between -1.0 and 1.0 (exclusive) could represent non-linearity in the individual outcomes that could reflect any number of selection strategies—some of which might involve using CLJs, and others might not (as will be supported by further analysis in the General Discussion; see Table 4 there). Thus, although the cutoffs (.50 and $-.50$) did not guarantee that the participants within each group used identical strategies, the choice was meant to be conservative. Moreover, the majority of participants had correlations either above .50 or below $-.50$

(see Figs. 1 and 2), so few participants were dropped from the analyses reported below. After forming the two groups, we compared them on a variety of measures, which are presented in Table 2 and discussed in their order of appearance below.

Perhaps participants used different selection strategies based on their study performance. For example, those who performed poorly during the study phase may have selected the easier categories for restudy, whereas those who performed well may have selected the remaining (difficult) categories. If so, performance during the study phase would be significantly lower for those who selected easier categories than for those who selected more difficult ones. However, the differences in performance during study (see the *Study Performance* row in Table 2) between the groups were relatively small and were only significant in Experiment 1b. Hence, performance differences prior to selection do not provide a sufficient explanation for the large differences in selection behavior.

Another possibility is that participants differed in how they judged their performance. For instance, Metcalfe and Finn (2008) found that when making study decisions, people were influenced more by their judgments than by their actual level of performance. Thus, some participants may have been less confident than others in their ability, and their reduced confidence may have driven them to select easier categories for restudy. If so, the mean CLJs would be lower for the easy than for the difficult group, but the differences in CLJs (see *Mean CLJ* in Table 2) between the groups were relatively small and trended in opposite directions between the two experiments.

We also compared the groups on the times to make their first three selections. The time to make the first selection could in part reflect how much time was devoted to planning (Dunlosky & Thiede, 2004). Thus, if one group planned more, they might have taken longer to make their first selection. The mean times to selection for both groups are presented in the rows *Time to 1st*, *2nd*, and *3rd Selection* in Table 2. For both

Table 2 Comparisons between easy and difficult category selectors in Experiments 1 and 2

	Experiment 1a		Experiment 1b		Experiment 2	
	Easy	Difficult	Easy	Difficult	Easy	Difficult
Sample size	47	48	38	48	34	25
Study performance	.24 (.01)	.27 (.01)	.23 (.01)*	.27 (.01)*	.22 (.02)	.25 (.02)
Mean CLJ	32.75 (2.06)	30.48 (1.90)	29.51 (2.09)	29.90 (2.14)	27.28 (1.99)*	37.84 (3.00)*
Time to 1st selection	4.64 (0.38)*	6.73 (0.48)*	3.65 (0.38)*	8.56 (1.46)*	3.96 (0.41)*	6.64 (0.67)*
Time to 2nd selection	3.92 (0.47)	4.05 (0.45)	3.20 (0.50)	4.53 (0.66)	3.28 (0.41)	4.2 (0.57)
Time to 3rd selection	3.63 (0.67)	3.19 (0.48)	3.21 (0.49)	3.52 (0.60)	3.38 (0.56)	4.25 (0.69)
Test performance	.28 (.02)	.31 (.02)	.28 (.02)*	.33 (.02)*	.26 (.02)	.27 (.02)
Study performance–CLJ ^a	.63 (.03)	.62 (.04)	.59 (.04)*	.71 (.02)*	.84 (.25)	.60 (.06)

Standard errors of the means are in parentheses. Participants with a gamma correlation between CLJs and selections at or above .5 were labeled as selecting easy categories for restudy (*Easy*). Those who had a gamma correlation at or below $-.5$ were labeled as selecting difficult categories for restudy (*Difficult*). * Significant difference between the easy and difficult groups at $p < .05$. ^a Correlation between each participant’s study performance and their CLJs for each category, averaged across participants.

experiments, the mean time to make the first selection was significantly greater for the difficult than for the easy group, but the mean times to make the second and third selections did not differ. These results indicate that the difficult group may have used more time before making their first selection to construct a plan about which categories to restudy.

For the last two measures presented in Table 2 (*Test Performance* and *Study Performance–CLJ*), we had no a priori predictions. Criterion test performance was greater (albeit not always significantly) for the difficult than for the easy group. Finally, we evaluated whether participants appeared to rely on performance during the study phase to make their CLJs by computing a within-participant correlation between study performance for each category and subsequent CLJs. Higher correlations would suggest that participants relied more on prior performance to make their CLJs (see the row labeled *Study Performance–CLJ*). Differences between the two groups were not consistent.

Category-learning judgment resolution Although the resolution (i.e., the relative accuracy) of CLJs was not relevant to the main aims of the present experiments, we briefly present the group means for archival purposes, especially given that few studies have investigated CLJs. We computed within-participant correlations between participants' CLJs and final test performance. Because our CLJ prompt had participants predict performance for identifying novel exemplars of each category, we correlated CLJs with correct performance for novel category exemplars presented in the final test. The means across participant correlations were .48 ($SEM = .02$) for Experiment 1a and .55 (.14) for Experiment 1b.

Experiment 2

In summary, the manipulations in Experiments 1a and 1b had a minor (and inconsistent) influence on how participants used CLJs to select categories for restudy. Instead, it appeared that some participants tended to select for restudy those categories that had been judged as relatively easy to learn (the positive correlations in Figs. 1 and 2), whereas others did the opposite (the negative correlations). These outcomes were unexpected given the prior literature, which typically reported mean judgment–selection correlations that were substantially and significantly different from zero (either positively or negatively, given the task constraints in the experiment; for a review, see Dunlosky & Ariel, 2011). Given that we discovered the bimodal distributions by exploring our data set, an important and major goal of Experiment 2 was to replicate (Kerr, 1998). To do so, we chose to use only a single selection group (the select-three group) with a larger sample size.

One explanation for the bimodal distributions could be that people strategically selected either the easier categories or the

more difficult ones. If people used these strategies explicitly, then they might report using them (e.g., “I decided to restudy the easier ones”), and these reports should relate to their actual selections (e.g., for the prior example, the participant should have a positive correlation between CLJs and selections). Alternatively, the bimodal distribution might result from participants randomly selecting categories or using another strategy that was not related to category difficulty. In such cases, we should observe no relationship between self-explanations and actual selections (Nisbett & Wilson, 1977). To evaluate these ideas in Experiment 2, participants made general retrospective reports of how they had selected categories for restudy. In particular, immediately after participants had selected the categories for restudy, they answered the following open-ended question: “How did you decide which bird families to restudy?”

Method

Participants A total of 85 participants completed the experiment for course credit. Four of the participants were removed due to a programming error, and one was removed for signing the wrong consent form, leaving 80 for analysis.

Materials The materials were the same as those used in Experiments 1a and 1b, but they also included a questionnaire. To construct the questionnaire in such a way to provide the most valid reports, we used the recommendations from Ericsson and Simon (1980). We began with an open-ended question that would be the least leading (“How did you decide which bird families to restudy?”). Any responses to this question that pertained to selecting easier or more difficult categories would be the most valid ones, because these particular strategies were not mentioned in the question prompt. However, this question could also lead to unrelated answers; hence, the questionnaire also included three more questions that queried participants about their specific strategies related to category difficulty: “Did you consider how well you learned each family, such as using the classification ratings (8%–100%) you made, to make your restudy decisions?,” “If so, how did you choose to use these ratings?,” and “Did you choose to restudy the bird families that you found easier to learn or the families you thought were difficult to learn?” Answers to the last three questions could be driven by how the questions were asked rather than what participants actually believed they had done when making selections. For the present analysis we focused only on the answers to the first question.³

³ The answers to the other questions were consistent with the answers to the first question. A majority (88%) reported using their CLJs to make restudy selections. The answers to the last question (which asked specifically whether participants had selected easy or difficult categories) were correlated with the answers to the first ($r = .57$).

Procedure The procedure was same as in Experiments 1a and 1b, except that all participants selected only three families for restudy; they were all presented with category names for selection simultaneously; and they were not shown their CLJs during the selection phase. After making their selections, participants filled out the questionnaire. Each question was presented on the computer individually. For the first three questions, participants typed their answer and then clicked a button to move on to the next question. For the last question, they selected one of the following options: *easy*, *difficult*, or *neither*. After filling out the questionnaire, participants completed the final test.

Results and discussion

Selection as a function of category-learning judgments We computed within-participant correlations between each participant's CLJs and selections. As in Experiments 1a and 1b, the average correlation across all participants did not differ significantly from zero: $M = .07$, $SEM = .09$, $t(75) = 0.84$, $p = .41$.

Frequency distributions As is evident from inspecting Fig. 3, the correlations between CLJs and restudy selections had a bimodal distribution, which replicated the outcomes reported in Experiments 1a and 1b.

Strategy questionnaire Immediately after category selection, participants answered the following open-ended question “How did you decide which bird families to restudy?” We used the outcomes from this question to evaluate whether participants reported strategically selecting easier or more difficult categories. We coded the reported strategies as selecting easier categories, selecting more difficult ones, using no

strategy, or using a different strategy. Two experimenters coded the data. Their interrater agreement was 88%, and any discrepancies were resolved through discussion.

Of the participants, 34 reported selecting easier categories, 35 reported selecting more difficult categories, and nine reported using no strategy or a strategy that was unrelated to category difficulty. We removed the latter nine participants from further analyses, because our main focus was on the difference between those who reported selecting easier versus more difficult categories. The reports were consistent with the correlations between CLJs and selections. The mean correlation between CLJs and selections for those who reported selecting the easier categories for restudy was significantly greater than zero: $M = .73$ ($SEM = .08$), $t(31) = 9.71$, $p < .001$. In contrast, the mean correlation between CLJs and selections for those who reported selecting difficult categories for restudy was significantly less than zero: $M = -.59$ ($.08$), $t(34) = 7.38$, $p < .001$. These outcomes support the conclusion that the bimodal distributions in Experiments 1 and 2 (Figs. 1, 2 and 3) reflect individual differences in people's strategic use of CLJs when selecting categories for restudy.

Selection group differences As in Experiments 1a and 1b, we separated participants into two groups: those who tended to select easier categories for restudy versus those who tended to select more difficult categories. We then compared the groups on several measures, as is shown in the two rightmost columns of Table 2. The groups did not differ significantly on study performance, although the results did trend in the expected direction. The mean CLJ was significantly higher for the difficult group than for the easy group, suggesting the difficult group was more confident. However, given that this measure

Fig. 3 Frequency distribution of correlations between category-learning judgments (CLJs) and selections for Experiment 2

had not been significant in Experiments 1a and 1b, we are hesitant to make strong conclusions regarding this outcome. As in Experiments 1a and 1b, the difficult group took significantly longer to make their first selection than the easy group, but the times to make the next two selections did not differ between groups. The two groups did not differ on test performance, $t(71) = 1.64$, $p = .11$, or in their correlations between study performance and CLJs, $t(67) = 0.58$, $p = .56$.

Category-learning judgment resolution The mean correlation between participants' CLJs and test performance for novel category exemplars was .60 ($SEM = .15$).

Experiment 3

In summary, the results from Experiments 1a, 1b, and 2 support the conclusion that people use CLJs to make study decisions (Figs. 1, 2 and 3), which illustrates that some participants selected those categories that they had judged to be relatively easy (i.e., higher CLJs), whereas others selected categories they had judged to be relatively difficult (lower CLJs). In contrast to expectations based on prior theory and evidence (e.g., Dunlosky & Ariel, 2011; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999), the various manipulations used in the prior experiments did not have a substantive impact on how participants made their restudy decisions. Instead, we consistently observed a bimodal distribution of CLJ–restudy selection correlations. Finally, exploratory analyses to identify any potential sources of differences in selection behavior did not yield consistent answers as to why the participants selected easier or more difficult categories for restudy.

In Experiment 3, we again (a) attempted to influence how participants used their CLJs to select categories for restudy and (b) measured other individual differences that might account for differences in selection behavior. For the former goal, we used a different version of our main manipulation from Experiments 1a and 1b. As before, one group of participants selected three categories for restudy, but the new group could select as many categories as they wanted (henceforth, the *free-choice group*). We included the free-choice group because such choice is similar to methods often used in the memory literature (for a review, see Dunlosky & Ariel, 2011; Son & Kornell, 2008). We expected the select-three group to replicate the other experiments, with participants tending to select for restudy categories judged to be either easier or more difficult. As compared to the select-three group, one prediction was that the free-choice group would be more likely to select categories that judged to be more difficult (i.e., given lower CLJs), simply because they have an opportunity to master more (and, hence, more difficult) categories. However, given that our previous manipulations had had little, if any, effect on participants' choices, our a priori expectation

was that this manipulation would also have a small impact on the relationship between CLJs and restudy selections.

As importantly, even if this manipulation (select-three vs. free-choice group) did not affect the CLJ–selection relationship, it would allow us to explore the degree to which another individual-difference measure—the number of categories selected—covaried with whether people tended to select categories they judged to be easier or more difficult. Beyond the number of categories selected, we collected the same measures as in the previous experiments (see Table 2) and also included new ones in order to evaluate specific hypotheses. First, one important difference between the two groups could be their general memory abilities. For instance, participants who have poorer memory could have more difficulty learning categories, and hence select easier categories for restudy. To address this possibility, we administered an associative memory task. Second, even if the groups did not differ in actual memory ability, they might differ in perceived ability (task self-efficacy) or perceived task difficulty. Participants who had lower perceived ability for the task (i.e., lower self-efficacy) might select the easier categories so as to try to perform well on at least some of them. By contrast, participants who had higher perceived ability for the task (i.e., higher self-efficacy) might select the more difficult categories, believing this strategy would maximize their performance. Third, concerning perceived task difficulty, some participants might have perceived the overall task as being more difficult and select the easier categories, but others might perceive the overall task as being easier and select the more difficult categories. Finally, another (non-mutually-exclusive) possibility is that individual differences in performance goals would be related to selection, with those adopting lower performance goals also selecting the easier categories. We measured these three constructs (perceived ability, perceived task difficulty, and goal) with a short questionnaire presented before the selection phase (described in detail below).

Method

Participants A total of 156 Kent State University undergraduates completed the experiment for course credit. Five of the participants were removed for not paying attention or leaving the experiment early, 21 were removed because they had participated in a similar experiment before, and 12 were removed due to computer errors. The final sample consisted of 118 participants.

Materials The materials were the same as in the previous experiments, but included an additional questionnaire and an associative memory task. The questionnaire consisted of three questions in addition to those presented in Experiment 2. Question 1 (perceived task difficulty) asked, “How easy was

it for you to learn to classify birds into their correct family?” Participants responded on a scale from 1 to 5, where 1 = *extremely difficult* and 5 = *extremely easy*. Question 2 (perceived ability) asked, “If you were tested right now on the bird categories that you just studied, how well do you think you would perform?” Participants responded on a percentage scale. Question 3 (goal) asked, “What is your goal for the final test? What percentage of all the birds do you want to correctly classify?”

The associative memory task consisted of 33 unrelated noun–noun pairs. Three pairs were used for practice, and the other 30 were used for the task. Participants first completed the practice trial. During practice, participants saw three unrelated noun–noun pairs, one pair at a time. Each pair was presented for 4 seconds. After all pairs had been presented, participants completed a cued-recall test. They typed the second word when prompted with the first, then pressed the Enter key. If they could not recall a word, they were asked to press the Enter key to move on to the next word. The actual task was similar, but participants studied 30 word pairs presented in the same random order for all participants. After study, participants were told to count backward by threes from a three-digit number for 2 minutes as a distractor task. Afterward they completed a cued-recall test on all 30 word pairs in a newly randomized order.

Procedure The procedure was similar to those of the previous experiments. After making CLJs, participants completed the questionnaire consisting of the questions from Experiment 2 and the three new questions. Then 55 of the participants were randomly assigned to the select-three group, and the remaining 63 participants were assigned to the free-choice group.

When the free-choice group was done making selections, they pressed a button at the bottom of the screen to continue. After the final test, all participants completed the associative memory task. The whole experiment lasted approximately 30 minutes.

Results and discussion

Selection as a function of category-learning judgments

Unlike in the previous experiments, the average within-participant correlation between participants’ CLJs and restudy selections was significantly less than zero: $M = -.41$, $SEM = .07$, $t(105) = 5.98$, $p < .001$. The mean correlations did not differ between the select-three group ($M = -.38$, $SEM = .09$) and the free-choice group ($M = -.45$, $SEM = .11$): $t(104) = 0.47$, $p = .64$, $d = 0.09$.

Frequency distributions As is shown in Fig. 4, the distribution was bimodal across all participants; however, the distribution was skewed toward negative values, which is consistent with the group mean being significantly negative.

Strategy questionnaire and associative memory task As in Experiment 2, we coded participants’ responses to the question “How did you decide which bird families to restudy?” as selecting *easier* categories, selecting *more difficult* ones, using *no strategy*, or using a *different strategy*. Twelve of the participants reported selecting easier categories, 92 reported selecting more difficult categories, and 13 reported using no strategy or a different strategy. We removed the latter group from our next analyses. The mean correlation between CLJs and selections for those who reported selecting the easier categories for restudy was significantly greater than zero: $M = .69$

Fig. 4 Frequency distribution of correlations between category-learning judgments (CLJs) and selections for Experiment 3

($SEM = .14$), $t(11) = 4.87$, $p < .001$. The mean correlation between CLJs and selections for those who reported selecting the more difficult categories for restudy was significantly less than zero: $M = -.61$ ($SEM = .06$), $t(87) = 10.21$, $p < .001$.

Across all participants, for perceived ability, participants gave a mean rating of 35% ($SEM = 2.09$), and for perceived task difficulty, they gave a mean rating of 1.94 ($SEM = 0.07$). Both ratings indicate that participants found the task difficult. On average, participants reported a goal of achieving 51% ($SEM = 2.27$) on the final test. The mean performance across all participants for the associative memory task was 19% ($SEM = 1.0$).

Selection group differences As in the prior experiments, we separated participants into easy and difficult groups and compared them in terms of potential individual differences. We also compared the groups on the five new measures: associative memory, perceived ability, perceived task difficulty, goal, and number of categories selected. The results are presented in Table 3. The groups did not differ significantly in study performance, but the results did trend in the expected direction. In contrast to the outcomes from the previous experiments, the time to first selection was not significantly greater for the difficult than for the easy group. All other outcomes were similar to those from the first three experiments.

Concerning the new measures, the overall performance on the associative memory task was low for all participants (Table 3, *Associative Memory* row), and although the results were in the expected direction, the difficult group did not significantly outperform the easy group. The groups also did

not differ significantly on perceived ability, perceived task difficulty, or their self-reported goal for the final test. Finally, for participants in the free-choice group (who could decide how many categories to restudy), the difficult group selected significantly more categories for restudy than did the easy group (see *Number Selected* in Table 3).

Category-learning judgment resolution The mean correlation between CLJs and performance for novel exemplars was .53 ($SEM = .02$).

Experiment 4

In Experiment 3, we tried again to manipulate the task constraints in a way that would change selection behavior, but our manipulation did not affect the CLJ–restudy selection relationship. In Experiment 4, we attempted to manipulate participants' goal for final test performance: They were instructed to achieve a goal of either 30% or 80% correct on the final test. We predicted that when given a lower goal, participants would tend to select the easier categories for restudy because they only had to master 30% of the birds (see, e.g., Thiede & Dunlosky, 1999). By contrast, participants with a higher goal of 80% were expected to select more of the difficult categories in an attempt to achieve the instructed goal. Of course, given our outcomes from prior experiments, we again expected this manipulation would have a minor impact.

Some of the results regarding individual differences from Experiment 3 were promising, and we wanted to replicate

Table 3 Comparisons between easy and difficult category selectors in Experiments 3 and 4

	Experiment 3		Experiment 4	
	Easy	Difficult	Easy	Difficult
Sample size	20	66	45	68
Study performance	.23 (.02)	.27 (.01)	.23 (.01)	.26 (.01)
Mean CLJ	39.17 (3.34)	32.54 (1.85)	31.65 (1.63)	30.67 (1.41)
Time to 1st selection	4.82 (0.48)	5.52 (0.41)	5.24 (0.36)	5.05 (0.32)
Time to 2nd selection	3.74 (0.52)*	5.78 (0.47)*	3.96 (0.34)	3.88 (0.42)
Time to 3rd selection	2.88 (0.46)	3.25 (0.40)	3.01 (0.36)	2.93 (0.44)
Test performance	.27 (.03)	.32 (.01)	.29 (.01)	.31 (.01)
Study performance–CLJ ^a	.73 (.05)	.66 (.02)	.58 (.05)	.64 (.02)
Associative memory	.17 (.03)	.19 (.02)	.15 (.02)*	.23 (.02)*
Perceived task difficulty	2.05 (0.18)	1.94 (0.09)	2.04 (0.10)	2.00 (0.09)
Perceived ability	37.15 (5.64)	36.12 (2.70)	38.96 (3.15)	37.51 (2.21)
Goal	49.75 (5.01)	52.32 (3.13)	N/A	N/A
Number selected ^b	3.90 (0.52)*	5.23 (0.31)*	4.44 (0.23)*	6.34 (0.24)*

* Significant difference between the easy and difficult groups at $p < .05$. ^a Correlation between each participant's study performance and CLJs for each category, averaged across participants. ^b For Experiment 3, this analysis was run only for participants in the free-choice group. Hence, the sample sizes are smaller for this group: easy = 12, difficult = 37.

them in Experiment 4. First, we wanted to replicate the small difference between the groups on the associative memory task. Second, we wanted to replicate the group differences in the number of categories selected for restudy, especially given that this analysis was exploratory. To do so, all participants could select as many categories as they wanted (as in the free-choice group of Exp. 3). We also included the questions about perceived ability and perceived task difficulty.

Method

Participants A total of 167 Kent State University undergraduates completed the experiment for course credit. One participant left the experiment early and was removed. A further 13 were removed because they had participated in a similar experiment before, and two were removed due to computer errors. The final sample consisted of 149 participants.

Materials The materials were the same as those used in the previous experiment, except that the question about participants' goal for the final test was removed because they were each assigned a goal.

Procedure The procedure was the same as in Experiment 3, but all participants could select as many categories for restudy as they wanted. Before making restudy selections, participants were given a goal for the final test. Of the participants, 61 were given a low performance goal, to correctly classify 30% of the birds, and the other 88 participants were given a high performance goal, to correctly classify 80%. Participants were told that if they achieved their goal, they would be entered in a drawing to win a \$30 Amazon gift card. After the experiment, five participants were randomly chosen out of all participants

(regardless of their final test performance) to receive a gift card.

Results and discussion

Selection as a function of category-learning judgments The average within-participant correlation between participants' CLJs and restudy selections did not significantly differ from zero: $M = -.14$, $SEM = .07$, $t(129) = 1.91$, $p = .06$. The groups also did not differ from each other: 30% goal, $M = -.16$, $SEM = .11$; 80% goal, $M = -.12$, $SEM = .10$; $t(128) = 0.23$, $p = .82$, $d = 0.04$.

Frequency distributions The frequency distribution is presented in Fig. 5. The distribution of participants' correlations between CLJs and selections was again bimodal.

Strategy questionnaire and associative memory task We coded participants' responses to the question "How did you decide which bird families to restudy?" as selecting *easier* categories, selecting *more difficult* ones, using *no strategy*, or using a *different strategy*. Of the participants, 31 reported selecting easier categories, 100 reported selecting more difficult categories, and 17 reported using no strategy or a different strategy. We removed the latter group from our next analyses. Consistent with these self-reports, the mean correlation between CLJs and selections for those who reported selecting the easier categories for restudy was significantly greater than zero ($M = .81$, $SEM = .07$), $t(28) = 11.97$, $p < .001$, and the mean correlation between CLJs and selections for those who reported selecting the more difficult categories for restudy was significantly less than zero ($M = -.48$, $SEM = .07$), $t(93) = 7.18$, $p < .001$.

Fig. 5 Frequency distribution of correlations between category-learning judgments (CLJs) and selections for Experiment 4

Across groups, participants gave a mean rating of 36% ($SEM = 1.68$) for perceived ability. They gave a mean rating of 1.98 ($SEM = 0.06$) for perceived task difficulty. The mean performance across all participants on the associative memory task was 19% ($SEM = 1.0$).

Selection group differences As is shown in Table 3 (two rightmost columns), analyses conditionalized on group differences largely replicated the outcomes from the previous experiments, with one exception being that the time to first selection did not differ significantly between the groups. Consistent with Experiment 3, the groups did not differ in their responses to perceived task difficulty and perceived ability. Most important, the difficult group significantly outperformed the easy group on the associative memory task and also selected significantly more categories for restudy than did the easy group.

Category-learning judgment resolution The mean correlation between CLJs and performance for novel exemplars was similar to those in the previous experiments, $M = .52$ ($SEM = .02$).

General discussion

Do people use their category-learning judgments to regulate their learning of natural categories? Outcomes from five experiments were consistent with the hypothesis that people do use their CLJs to decide which categories to restudy. Nevertheless, some of the outcomes were unexpected: First, manipulations that were expected to influence how people used CLJs had small and inconsistent effects. More intriguing, people used different strategic approaches when using CLJs to make restudy selections, which resulted in extreme individual differences in selection behavior (Figs. 1, 2, 3, 4 and 5). We first briefly discuss the minor impact of the manipulations, and then use the remainder of the General Discussion to discuss the individual differences in the strategic use of CLJs, why they may have occurred, and whether the different strategic approaches could both be effective.

On the basis of the agenda-based regulation framework (Ariel et al., 2009), we expected participants to select the more difficult unlearned categories when they could study many categories (the select-nine group in Exps. 1a and 1b), and to select the easier categories when they could study only a few (the select-three group in Exps. 1a and 1b). The two groups did not consistently and significantly differ in how they made restudy selections (Exp. 1a, $d = 0.70$; Exp. 1b, $d = 0.28$), but the trends were in the expected direction, so in subsequent experiments we attempted to influence selection behavior by using manipulations that could potentially influence how people used CLJs when making restudy selections. However,

these manipulations did not influence selection behavior, either (Exp. 3: select three vs. free choice, $d = 0.09$; Exp. 4: high vs. low goal, $d = 0.04$).

Why did these manipulations not consistently influence the relationship between CLJs and selections, as expected? One possibility is that the impact of the manipulations was constrained by task difficulty along with individual differences in selection behavior. Consider the outcomes from Experiment 4. Although some participants were instructed to achieve a high goal (80%), everyone perceived the task as being rather difficult (see Table 3); in fact, the task was objectively difficult across all experiments. In such cases, given that participants apparently had different approaches to using CLJs to make restudy decisions (Fig. 5), they may have disregarded the instructions and instead used the strategy that they believed would be the most effective one. Put differently, despite being given a high learning goal in Experiment 4, many participants may not have believed that they could achieve it, so they elected to use a strategy that would potentially increase their chances of boosting their performance. Albeit this account is speculative, it could explain why the manipulations did (somewhat) influence how participants used CLJs in Experiments 1a and 1b, because in those earlier experiments, participants were *forced* to select either three or nine categories. Future research will be needed to more fully explore when these kinds of manipulation will (vs. will not) impact how people use CLJs.

Individual differences in the strategic use of CLJs: Why do they occur?

Given that the impact of the aforementioned manipulations was inconsistent in the present research, we do not consider them further, and instead turn our focus to a consistent and novel outcome from all experiments: Some participants tended to select the categories they had judged as being more well learned, and others selected the categories they had judged as being less well learned. Because this outcome was unexpected, we used two other analytic approaches to evaluate individual differences in the relationship between CLJs and selection behavior (for details, see the Appendix), and both analyses supported the same conclusions about the role of individual differences. Also, in three follow-up experiments (Exps. 2, 3, and 4), we had participants report how they had selected categories for restudy. Their reports were consistent with how they made their selections, providing converging evidence that CLJs are related to category selections, and that individual differences exist in how people use CLJs to make those selections.

Why do these individual differences occur in the use of CLJs? According to the agenda-based regulation framework, monitoring and control processes interact directly with one's memory and belief systems (Dunlosky & Ariel, 2011), and

hence can be influenced by any number of individual differences (see also Ariel, Dunlosky, & Bailey, 2009; Nelson & Narens, 1990; Winne & Hadwin, 1998). For example, people's perceived ability (how well they believe they can perform a task) could interact with control decisions, in that people with greater perceived ability might believe they can excel at the task, and so try to learn the more difficult categories. Another possibility is that the integrity of the memory system itself (e.g., how well one can develop associations) is relevant, with those who have more difficulties making associations also selecting those categories that they had judged easier to learn. To provide a preliminary exploration of the contributions of these and other factors in Experiments 3 and 4, we measured several theoretical constructs that could plausibly explain the individual differences (for details, see the introduction to Exp. 3). In most cases we did not find the expected relationship; that is, the two groups did not differ on many of the measures (Tables 2 and 3).

Nevertheless, Experiments 3 and 4 included two measures that do provide initial evidence for how the two groups may differ. First, across both experiments, more categories were selected by those in the difficult group than by those in the easy group. Selecting more categories may indicate greater motivation to perform well on the final test. If so, one might also expect that participants who were more motivated would also have a higher goal for final test performance, but the groups did not differ in their reported goals for final test performance (Exp. 3). Thus, the evidence for the role of motivation is mixed. Second, the integrity of the memory system could contribute to individual differences, because people who have difficulties making associations (or poorer memory in general) might select easier categories in an attempt to compensate. Consistent with this possibility, in Experiments 3 and 4, associative memory performance was greater for those in the difficult than for those in the easy group (Table 3; albeit this trend was small and not statistically

significant in Exp. 3). Beyond people's ability to learn, individual differences in selection behavior could be related to their perceived ability to perform well on the task, with those having lower self-efficacy selecting easier categories and those with higher self-efficacy selecting more difficult ones. In contrast to this possibility, however, people's perceived abilities did not differ between the two groups in either Experiment 3 or 4.

Alternatively, people may have used CLJs differently to make restudy decisions because they believed that their approach was an effective one. To evaluate this hypothesis, we examined participants' reported strategy use from the questionnaire in Experiments 2, 3, and 4 (these results are also presented in the Results section for each of these experiments). We coded the strategy reports as selecting *easy* categories, selecting *difficult* categories, using *other strategies*, or using *no strategy* and combined the results across experiments. The results are presented in Table 4. For each kind of strategy report, we also present representative responses, the proportion of participants who reported each strategy, and the mean correlation between CLJs and selection. As we emphasized above, participants' reports of how they used CLJs are consistent with their behavior (e.g., those who said they had restudied the easier categories on average had a positive correlation between CLJs and selection). We did not ask participants to explain why they selected the easier (or more difficult) categories for restudy (for a rationale, see Nisbett & Wilson, 1977). Even so, some people spontaneously indicated that they had used a strategy that they believed was effective. For instance, as is shown in the *Example Responses* column (Table 4, *Easy* group), some participants reported that if they could memorize the easy categories, then they might be able to figure out the more difficult ones, which is consistent with a region-of-proximal-learning strategy (Metcalf, 2009; Son & Metcalf, 2000). Others indicated selecting categories that

Table 4 Participants' responses to the question "How did you decide which bird families to restudy?"

Response Categorization	Example Responses	Percentage	Mean Correlation
Easy	"I felt the three I chose were the ones I'd have the best chance at remembering." "Because those were the easiest to understand, so if I can memorize them I can possibly figure out the rest."	23%	.76 (.41)
Difficult	"I picked the ones I felt least confident with when I was originally studying them." "The ones that I seemed to get incorrect most often."	65%	-.55 (.59)
No Strategy	"Swallow" ^a "I didn't know any of them, so I picked randomly."	3%	.43 (.40)
Other	"I knew that if I picked one family I already knew fairly well (Orioles), then it would aide me in distinguishing between the other two families I selected..." "I was going to choose the ones I couldn't distinguish very well, then decided I might as well restudy all of them."	8%	.26 (.67)

Example responses are sample reports from each of the four categorizations. The mean correlation is the correlation between CLJs and selections for each categorization. Standard deviations are in parentheses. Correlations were calculated for a subset of participants, because correlations could not be calculated for everyone. The sample sizes used to calculate correlations for each categorization are as follows, in the order presented in the table: 75, 217, 9, and 11. ^a Some participants only listed the bird categories they had selected.

they would have the best chance of remembering. Reports about strategy effectiveness were not explicit in many responses (because we did not ask them to explain their selections), but along with the consistent reports (88%) of selecting easy versus difficult categories, they do support the general conclusion that people strategically used the CLJs. Challenges for future research include revealing why people differ in what they believe will be an effective strategy for this task, as well as the degrees to which the different strategies are equally effective. We consider the latter issue next.

Are both selection strategies effective?

One issue that is of general interest in self-regulation research is the extent to which the strategies that people use to allocate study improve their performance (Kimball, Smith, & Muntean, 2012; Kornell & Metcalfe, 2006; Metcalfe, 2009; Metcalfe & Finn, 2013; Mulligan & Peterson, 2014; Nelson, Dunlosky, Graf, & Narens, 1994; Rhodes, Sitzman, & Rowland, 2013; Son, 2010). This research investigated people's selection of items (e.g., single words or paired associates) for a memory test. In most of this research, how people allocated their study time appeared to improve their subsequent learning. In the present study, the issue was whether both strategies (selecting easier vs. more difficult categories) might have been effective for the subset of participants who chose to use them. Investigating this issue would require different methods (e.g., the honor–dishonor method of Kornell & Metcalfe, 2006) from the ones used in the present study, but recent evidence from Tullis and Benjamin (2011) suggests that both strategies may not be equally effective. Tullis and Benjamin had participants self-pace their study of individual words; some participants spent more time studying the normatively difficult words, whereas others spent more time studying the normatively easy words. The participants who spent more time studying the more difficult words performed better at final test than did those who spent more time studying the easier words, suggesting that spending more time on difficult items is a better strategy for learning word lists.

Although the data from Tullis and Benjamin (2011) suggest that both strategies are not equally effective for memory regulation (see also Dunlosky & Connor, 1997; Thiede, 1999), Tullis and Benjamin's methods differed substantially from those used here, and hence their results may not transfer to the present task. In particular, self-paced study of individual items (their task) and selection of categories (present task) differ with respect to task (pacing study vs. selecting for study), to content (words vs. birds), and to task goals (memorize vs. categorize), and any of these differences could contribute to different performance outcomes. For instance, in the present task there were only 12 categories, and participants typically selected bird categories from an array, which has been shown to promote planning that helps people achieve

their task goals (e.g., Dunlosky & Thiede, 2004). Thus, although this account is speculative, perhaps the present task promoted planning that would better align each participant's abilities with the strategy that would be most effective. We leave the resolution of this issue for future research.

Closing remarks

The results from the present experiments are consistent with the hypothesis that many people consider how well they have learned different categories when deciding which ones to re-study. Moreover, individual differences in how people used CLJs to select categories for restudy were demonstrated in all five experiments. Given the novelty of these individual differences in selection behavior within a currently understudied context (i.e., self-regulated learning of categories), the results pose important issues for future research. These include discovering why individual differences in restudy selection occurred and whether the different selection behaviors were equally effective for those who chose to use them.

Given that the present experiments represent the first to explore the relationship between people's CLJs and the selection of categories for restudy, many other avenues remain open for future research. Consider two more: First, having participants make CLJs may influence how they make their restudy selections. For example, Mitchum, Kelley, and Fox (2016) demonstrated that participants were more likely to allocate study time to easier paired associates (vs. more difficult ones) when they made judgments of learning (JOLs) than for when they did not make JOLs. Mitchum et al. speculated that participants changed their learning goal when making JOLs, which in turn influenced how they paced their study. Discovering whether making CLJs has a similar reactive effect (on goal development and selection) will be an important area for theoretical and applied research.

Second, individual differences in restudy selection (i.e., as demonstrated in the present experiments, Figs. 1, 2, 3, 4 and 5) may not extend to learning other categories. For instance, the task of learning the natural categories in the present research was normatively difficult, partly because bird exemplars do not have features that define the category to which they belong. The overall task difficulty may have enticed some people to focus on easier categories. In contrast, when people are learning artificial categories in which the exemplars for each category are defined by a distinctive feature, the strategy of focusing on the easiest categories may not be viewed as necessary. In such cases, most participants may select the more difficult categories for restudy, minimizing individual differences in how CLJs are used to make restudy decisions. Future research could investigate this issue by using the present method with other natural categories (e.g., fish or trees) and artificial categories (e.g., fribbles; see Barry, Griffith, De Rossi, & Hermans, 2014).

In summary, people intentionally try to learn categories in many contexts—including medical doctors learning to categorize diseases, park rangers learning categories of plants and trees while training on the job, and even would-be bird watchers learning bird families before venturing into the wild, to name just a few. Given that research on self-regulated category learning is in its infancy (for another approach not involving CLJs, see Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013), further understanding and improving people's self-regulation of learning categories will be a key agenda for future investigation that promises to have broad implications for theory and a variety of applications.

Appendix

To calculate the correlation between CLJs and selections, we computed gamma correlations, which can be influenced by tied values. A gamma correlation is based on the number of concordances and discordances that occur for a participant's CLJs and selections across categories, with ties being dropped from the computation (for details, see Nelson, 1984). Hence, if few comparisons (concordances and discordances) remain for a participant (i.e., if most comparisons yield tied values), then the correlation could spuriously be drawn toward extreme values (−1 and 1). In the present case, for each participant, the maximum number of concordances and discordances that could arise when making comparisons between the CLJs and selections across the 12 categories is 27. Across all experiments, the total number of comparisons was close to the maximum (median = 25.0, interquartile range: 21.0–27.0). Thus, a preponderance of ties could not have led to biased estimates of the relationship between CLJs and selections as measured with the gamma correlation. Second, we examined each participant's bias to select easy (or difficult) categories using a different measure that would not be influenced by tied comparisons. Namely, for each participant, we computed the mean CLJ for the categories selected and for those not selected, and then computed the difference between the two values. Across the participants from all experiments, this measure was highly correlated with the gamma correlations ($r = .88$).

We also compared the mean CLJs for the selected and nonselected categories (see Table 5) within conditions. As expected, these too supported the same conclusions. That is, most differences were not significant (as per the nonsignificant correlations between CLJs and selections), and any significant differences were consistent with the results of the correlational analyses. In Experiment 1a, for the select-three group, the mean CLJs were greater for selected than for nonselected categories, whereas for the select-nine group, the mean CLJs tended to be lower

Table 5 Mean category-learning judgments (CLJs) for selected and nonselected categories as a function of condition

	Mean CLJ Selected	Mean CLJ Nonselected	Selected vs. Nonselected
Exp. 1a			
Select 3	42.11 (3.88) ^a	28.74 (2.29)	$t(43) = 3.24, p < .01$
Select 9	26.91 (2.11)	39.47 (3.60)	$t(45) = 3.34, p < .01$
Exp. 1b			
Select 3	26.02 (2.64)	26.62 (1.61)	$t(59) = 0.20, p = .84$
Select 9	31.48 (2.34)	35.68 (3.76)	$t(45) = 1.10, p = .28$
Exp. 2			
	34.12 (2.60)	29.10 (1.91)	$t(79) = 1.60, p = .11$
Exp. 3			
Select 3	24.96 (2.71)	33.21 (2.03)	$t(54) = 2.67, p < .05$
Free Choice	27.66 (2.93)	42.71 (3.38)	$t(52) = 3.34, p < .01$
Exp. 4			
30%	31.30 (3.04)	34.79 (2.52)	$t(51) = 0.79, p = .44$
80%	31.76 (2.42)	35.24 (2.18)	$t(79) = 0.94, p = .35$

^a Standard errors of the means are in parentheses.

for the selected than for the nonselected categories (which are consistent with the positive and negative CLJ–selection correlations for the select-three and select-nine groups, respectively, reported in Table 1). The same consistency (between the conditional analyses of mean CLJs in Table 5 and the CLJ–selection correlations) is evident for all experiments.

References

- Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, 138, 432–447. doi:10.1037/a0015928
- Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the fribbles: Novel stimuli for use within behavioral research. *Frontiers in Psychology*, 12, 103. doi:10.3389/fpsyg.2014.00103
- DeSoto, K. A., & Votta, C. M. (2016). Psychology data on the effects of study schedules on category-member classification. *Journal of Open Psychology Data*, 4(26), 1–4. doi:10.5334/jopd.26
- Doyle, M. E., & Hourihan, K. L. (2016). Metacognitive monitoring during category learning: How success affects future behaviour. *Memory*, 24, 1–11. doi:10.1080/09658211.2015.1086805
- Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 54, pp. 103–140). San Diego: Elsevier.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, 25, 691–700. doi:10.3758/BF03211311
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, 98, 37–56. doi:10.1016/S0001-6918(97)00051-6

- Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory & Cognition*, 32, 779–788. doi:10.3758/BF03195868
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251. doi:10.1037/0033-295X.87.215
- Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics? *Contemporary Educational Psychology*, 49, 80–90. doi:10.1080/09658211.2016.1152378
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1441–1451. doi:10.1037/a0020636
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196–217.
- Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 923–954. doi:10.1037/a0026936
- 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. doi:10.1037/0278-7393.32.3.609
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18, 159–163. doi:10.1111/j.1467-8721.2009.01628.x
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179. doi:10.3758/PBR.15.1.174
- Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, 8, 19–46. doi:10.1007/s11409-013-9094-7
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463–477. doi:10.1016/j.jml.2004.12.001
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145, 200–219. doi:10.1037/a0039923
- Mulligan, N. W., & Peterson, D. J. (2014). The spacing effect and metacognitive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 306–311. doi:10.1037/a0038866
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133. doi:10.1037/0033-2909.95.1.109
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5, 207–213. doi:10.1111/j.1467-9280.1994.tb00502.x
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). San Diego: Academic Press.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259. doi:10.1037/0033-295X.84.3.231
- Rhodes, M. G., Sitzman, D. M., & Rowland, C. A. (2013). Monitoring and control of learning own-race and other-race faces. *Applied Cognitive Psychology*, 27, 553–563. doi:10.1002/acp.2948
- Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 255–262. doi:10.1037/a0017892
- Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 333–351). New York: Psychology Press.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 204–221. doi:10.1037/0278-7393.26.1.204
- 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. doi:10.1016/j.actpsy.2014.11.011
- 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. doi:10.3758/s13423-012-0319-6
- Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, 6, 662–667. doi:10.3758/BF03212976
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024–1037.
- Thomas, R. C., Finn, B., & Jacoby, L. L. (2016). Prior experience shapes metacognitive judgments at the category level: The role of testing and category difficulty. *Metacognition and Learning*, 11, 1–18. doi:10.1007/s11409-015-9144-4
- Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, 64, 109–118. doi:10.1016/j.jml.2010.11.002
- Wahlheim, C. N., & DeSoto, K. A. (2017). Study preferences for exemplar variability in self-regulated category learning. *Memory*, 25, 231–243. doi:10.1080/09658211.2016.1152378
- 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. doi:10.3758/s13421-010-0063-y
- 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. doi:10.3758/s13421-011-0180-2
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale: Erlbaum.
- Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, 145, 918–933. doi:10.1037/xge0000177
- Zauhar, V., Bajšanski, I., & Domijan, D. (2016). Concurrent dynamics of category learning and metacognitive judgments. *Frontiers in Psychology*, 7(1473), 1–11. doi:10.3389/fpsyg.2016.01473