Contents lists available at ScienceDirect

Cognition

Original Articles

Optimal sequencing during category learning: Testing a dual-learning systems perspective

Sharon M. Noh [a,*], Veronica X. Yan [b], Robert A. Bjork [b], W. Todd Maddox [a]

[a] University of Texas, Austin, United States
[b] University of California, Los Angeles, United States

A B S T R A C T

Recent studies demonstrate that interleaving the exemplars of different categories, rather than blocking exemplars by category, can enhance inductive learning—the ability to categorize new exemplars—presumably because interleaving affords discriminative contrasts between exemplars from different categories. Consistent with this view, other studies have demonstrated that decreasing between-category similarity and increasing within-category variability can eliminate or even reverse the interleaving benefit. We tested another hypothesis, one based on the dual-learning systems framework—namely, that the optimal schedule for learning categories should depend on an interaction of the cognitive system that mediates learning and the structure of the particular category being learned. Blocking should enhance rule-based category learning, which is mediated by explicit, hypothesis-testing processes, whereas interleaving should enhance information-integration category learning, which is mediated by an implicit, procedural-based learning system. Consistent with this view, we found a crossover interaction between schedule (blocked vs. interleaved) and category structure (rule-based vs. information-integration).

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

When learning new categories, how should the study of category exemplars be sequenced so that learners can accurately classify new exemplars on a later test? When an art student, for example, must learn to recognize the styles of different artists so as to be able to identify the artist responsible for a never-before-seen painting, should he or she study examples of artists' paintings one artist at a time, or should the paintings by the different artists be intermixed? Recent findings suggest that in this case, and in the inductive learning of other naturalistic categories, such as butterflies and birds, interleaving exemplars of different categories yields better category learning than does blocking exemplars by category (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim, Dunlosky, & Jacoby, 2011). More recent work using artificial stimuli suggests, however, that interleaving is only superior when between-category discriminability is low, and that blocking is superior when between-category discriminability is high (e.g., Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013). The important implication of these

studies is that there may be no single "optimal" method of sequencing, but rather, the optimal method may depend on various factors (e.g., the nature of the to-be-learned categories).

Although category discriminability can play an important role in determining whether interleaved or blocked study enhances category learning, we argue that another, yet unexplored, factor may be important: the learning system that mediates performance. An extensive body of behavioral, neuropsychological, and neuroscience literature suggests that optimal learning of different category structures is mediated by at least two neurobiologically grounded and competing learning systems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2011; Maddox & Filoteo, 2005; Nomura & Reber, 2008). One is a frontally mediated hypothesis-testing system that relies on working memory and executive attention to develop and test verbalizable rules that are used to optimally solve rule-based (RB) categories. The second is a striatally mediated procedural-based learning system that does not rely on working memory and executive attention but, instead, learns non-verbalizable stimulus-response mappings that are used to solve information-integration (II) categories. These two systems compete and previous research show that there is an initial bias toward the hypothesis-testing system, with control being passed to the procedural-based learning system only when the category structure warrants (e.g., with information-integration categories;

* Corresponding author at: 108 E. Dean Keeton Stop A8000, Department of Psychology, University of Texas at Austin, Austin, TX 78712-1043, United States.
E-mail address: smnoh@utexas.edu (S.M. Noh).

Ashby & Maddox, 2011; Ashby et al., 1998; Maddox & Ashby, 2004). Dual-learning-systems research suggests that learning in each system is optimized under different training conditions. For instance, rule-based category learning is optimized when full feedback is provided (e.g., "Wrong, that was a B") whereas information-integration category learning is optimized when immediate, minimal feedback is provided (e.g., "Wrong"; Maddox, Love, Glass, & Filoteo, 2008).

We hypothesize that the optimal schedules for category learning are also dependent on the underlying category structure. In the current study, we tested this hypothesis directly. With respect to rule-based categories, blocking exemplars by category should allow individuals to more easily generate, test, and adjust their working hypotheses, particularly when there is a relatively demanding working memory load. To introduce a working memory load, we used a four-category variant of the rule-based and information-integration learning structures (from Maddox, Filoteo, Hejl, & Ing, 2004), rather than the more typical two-category learning variant found in many dual-learning systems studies. An interleaved schedule, on the other hand, would hurt rule-based learning by introducing a more demanding working memory load, as individuals would have to generate and test multiple rules for each category simultaneously. While interleaving would allow learners to compare exemplars that do and do not fit into a given category, the working memory load involved in holding multiple dimensions in mind for multiple categories would make using rule-based hypothesis testing difficult. We predict that blocked study should better should facilitate rule-based category learning than interleaved study in our experiments. Following the same reasoning, we also hypothesize that interleaved study should be beneficial for information-integration category learning because it discourages the use of rule-based strategies and speeds the transition to the procedural based learning system.

2. Experiment 1

2.1. Method

2.1.1. Participants and design

One hundred and thirty-two participants (mean age = 30.0, age range = 19–57, 71 females) were recruited from Amazon Mechanical Turk (MTurk) and paid $1.00 for their participation. Category structure (rule-based vs. information-integration) and study schedule (blocked vs. interleaved) were manipulated in a 2×2 between-subjects design. An *a priori* power analysis determined that for a medium effect size ($f = 0.25$), we would need 32 participants per condition to reach a power of 0.80.

2.1.2. Materials

The four-category rule-based and information-integration category structures are displayed in Fig. 1. Each stimulus was comprised of a line of varying length and orientation at a fixed distance from center (that varied in position) on the computer screen. The stimuli were constructed from three continuous-valued dimensions: line orientation (between 0 and 90°), line length (0–200 pixels), and position (between 0 and 100° offset from fixation). Each dimension has eight values at equal intervals, but only line length and line orientation values defined category membership. Each of the eight line length values were paired with each of the 8 line orientation values, for a total of 64 unique lines of varying length and orientation. These 64 lines were randomly paired with one of 8 different positions, so that each unique line could be shown in one of 8 positions on the screen. In the rule-based condition, the stimulus space was divided into four categories using decision bounds that were verbalizable (e.g., "all

members of category A contain a short, steep line"). To generate the information-integration condition, the category boundaries and stimuli from the rule-based condition were rotated 45° so that no simple verbalizable rule could define category membership. This transformation allows us to both differentiate rule-based and information-integration category-learning strategies while keeping the category structures and stimulus distributions mathematically equivalent.

2.1.3. Procedure

Participants were asked to learn to distinguish exemplars from four different categories. A cover story was provided, suggesting that these were images generated by four different robots and the task was to learn each robot's way of generating images. During the study phase, participants observed each of the 64 images (constructed from the factorial combination of all 8 line lengths with all 8 line orientations) once with a randomly selected (but without replacement) position. Each item was presented with the appropriate category label (A, B, C, or D) for 3.5 s each. In the blocked condition, participants saw the 16 exemplars from one category before moving on to the next, whereas in the interleaved condition, all 64 exemplars were presented in a randomized order. Fig. 2 shows examples of the sequencing and stimuli used in the study phase. Following this passive study phase, participants moved on to the test phase, where they were shown the same 64 length-orientation pairings. The test stimuli were randomly presented, and following each stimulus presentation, participants were asked to select what they believed to be the appropriate category label by clicking on one of four buttons (labeled A, B, C, and D) arranged horizontally below each stimulus display. This final test was self-paced and without feedback.

2.2. Results and discussion

2.2.1. Classification performance

Average final test performance for each condition is presented in Fig. 3. A 2×2 between-subjects ANOVA revealed a main effect of category structure such that accuracy was higher for information-integration category structures ($M = 0.58$, $SD = 0.17$), relative to rule-based structures ($M = 0.51$, $SD = 0.19$), $F(1,128) = 5.36$, $MSE = 0.03$, $p = 0.022$, $\eta_p^2 = 0.04$. There was no significant main effect of schedule, $F(1,128) = 0.30$, $MSE = 0.03$, $p > 0.20$. There was, however, a significant interaction, $F(1,128) = 5.34$, $MSE = 0.03$, $p = 0.022$, $\eta_p^2 = 0.04$. Post-hoc t-tests revealed that for rule-based categories, accuracy following blocked study ($M = 0.55$, $SD = 0.19$) was marginally higher than accuracy following interleaved study ($M = 0.46$, $SD = 0.19$), $t(66) = 1.96$, $p = 0.055$, $d = 0.47$. The pattern was reversed, however, for information-integration categories: Accuracy following interleaved study ($M = 0.61$, $SD = 0.17$) was higher than accuracy following blocked study ($M = 0.55$, $SD = 0.17$), but this difference was not found to be significant $t(62) = 1.30$, $p = 0.20$, $d = 0.33$.

2.2.2. Model fits

The accuracy-based analyses suggest that blocking enhances RB learning, whereas interleaving helps II learning. We hypothesized that this effect would occur because blocking may facilitate hypothesis-testing and the rule-discovery process, whereas interleaving may discourage rule use (perhaps by introducing a working memory load). To examine this possibility, we fit a number of different decision bound models (e.g., Ashby & Gott, 1988; Maddox & Ashby, 1993) to the data from each individual participant in order to understand the kind of strategy each participant used to classify the stimuli. For each of the four experimental conditions, the relevant models were fit separately to the data from the 64-trial test block.
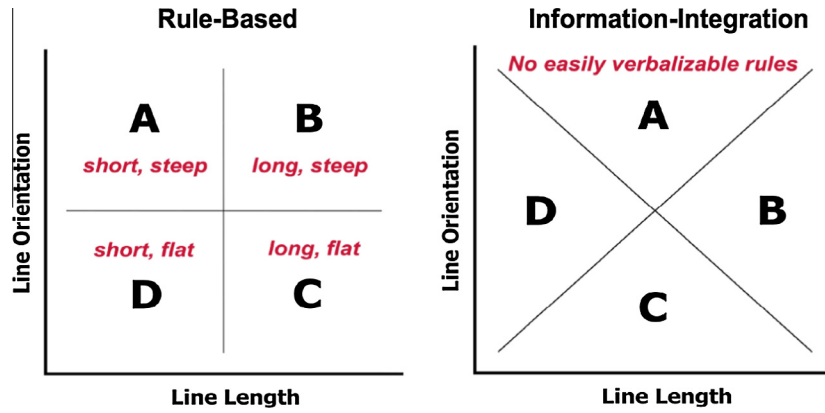
Fig. 1. The category structures of the RB (left) and II (right) categories. This figure displays the two relevant dimensions (line length and line orientation). In Experiment 1, there was one irrelevant dimension (position) and in Experiment 2, there were two irrelevant dimensions (ellipse length and position). These irrelevant dimension values varied randomly across stimuli, and are not illustrated here.
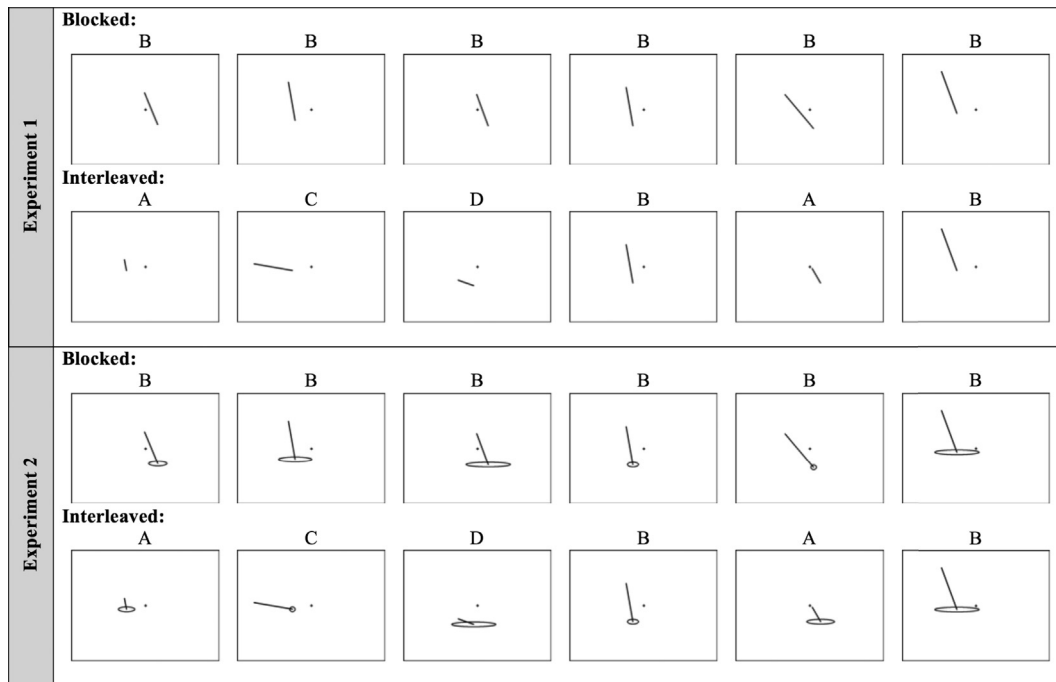

Fig. 2. Sample stimuli and sequencing for each condition. All exemplars were presented sequentially, one by one. In the blocked condition, all exemplars from the one category were presented before moving onto exemplars of the next category. In the interleaved condition, exemplars from all four categories were intermixed.
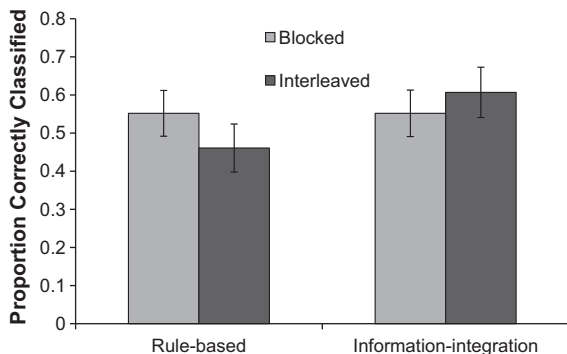

Fig. 3. Classification accuracy results, by schedule and category structure. Error bars represent 95% confidence intervals.

The model parameters were estimated using maximum likelihood (Ashby, 1992) and the goodness-of-fit statistic was computed using the Bayesian information criterion (BIC; Schwarz, 1978). The BIC is defined as $BIC = r \ln N - 2 \ln L$, where r is the number of free parameters, N is the sample size, and L is the likelihood of the model given the data. The BIC statistic penalizes models for extra free parameters. To determine the best-fitting model within a group of competing models, the BIC statistic is computed for each model, and the model with the smallest BIC value is reported as the best fitting model. Five different types of models were fit to each participant's responses: models that assumed an RB strategy using only one of the two relevant features (unidimensional rule use based on line length or line orientation), models that assumed an RB strategy using both relevant dimensions (conjunctive rule using both line length and orientation), models that assumed an II strategy, and models that assumed random guessing (for a detailed description of these models, see Maddox et al., 2004). The model

Table 1
Number (percentage) of participants best fit by each type of model for blocked and interleaved study schedules in the rule-based and information-information category structure conditions in Experiment 1.

Best-fit model	Rule-based structure		Information-integration structure	
	Blocked	Interleaved	Blocked	Interleaved
Information-integration	9 (25.0%)	11 (34.4%)	25 (71.4%)	24 (82.8%)
Conjunctive rule	21 (58.3%)	15 (46.9%)	7 (20.0%)	0 (0.0%)
Unidimensional length	4 (11.1%)	3 (9.4%)	0 (0.0%)	2 (6.9%)
Unidimensional orientation	1 (11.1%)	0 (9.4%)	0 (0.0%)	1 (6.9%)
Random responder	1 (2.8%)	3 (9.4%)	3 (8.6%)	2 (6.9%)
Average percentage of responses accounted for by best-fit model	68.0%	59.1%	63.8%	68.1%

fitting results are shown in Table 1. In the rule-based category structure conditions, the best-fitting model for each participant captured, on average, 68.0% and 59.1% of the variance in the responses made in the blocked and interleaved conditions, respectively. In the information-integration conditions, the best-fitting model for each participant correctly accounted for, on average, 68.1% and 63.8% of the variance in responses made in the blocked and interleaved conditions, respectively.

Consistent with our hypothesis, we found that blocked study led to proportionally more rule-based strategy users relative to interleaved study. This pattern held true, both when we compare the proportion of blocked and interleaved study condition participants who use any rule-based strategy (i.e., unidimensional length, unidimensional orientation, or conjunctive rule) and we restrict the comparison to only those who specifically used the conjunctive rule-based strategy. With the rule-based category structure condition, we found that 58.3% of participants in the blocked schedule were best fit by models assuming the conjunctive rule-based strategy (which is the optimal strategy for rule-based categories), relative to 46.9% in the interleaved schedule condition.

As predicted, interleaved study led to proportionally more information-information based strategy use relative to blocked study. With information-integration category structures, 82.7% of participants in the interleaved schedule condition were best fit by models assuming an information-integration strategy (which is the optimal strategy for these information-integration categories), relative to 71.4% of participants in the blocked schedule condition.

3. Experiment 2

Experiment 1 showed that blocking facilitated conjunctive rule use and enhanced the learning of rule-based categories, whereas interleaving discouraged rule-based strategies and enhanced the learning of information-integration categories. Computational models were applied to gain these insights into the strategies people may be using during the categorization process. To facilitate computational modeling in Experiment 1, we used highly controlled and relatively simple stimuli. Given that most categorization problems in the real world involve perceptually rich stimuli with many dimensions that are irrelevant to category membership, in Experiment 2, we replicated the design of Experiment 1 using more complex stimuli. To achieve this aim, we added a second irrelevant dimension to each stimuli—an ellipse with fixed height and varying length.

3.1. Method

3.1.1. Participants and design

One hundred and ninety-two participants (mean age = 34.44, age range = 18–59, 106 females) were recruited from MTurk and paid $2.50 for their participation. Category structure (rule-based vs. information-integration) and schedule (blocked vs. interleaved) were manipulated in a 2 × 2 between-subjects design. An a priori power analysis determined that for a medium effect size ($f = 0.25$), we would need 32 participants per condition to reach a power of 0.80. As we might expect more noise from the additional irrelevant dimension and therefore, a smaller effect size, we aimed instead for 50 participants per condition (based on $f = 0.20$).

3.1.2. Materials

The four-category rule-based and information-integration category structures were very similar to that of Experiment 1, with one additional, category-irrelevant dimension. In addition to line length, line orientation and position, each stimulus also consisted of an ellipse of varying length (from which the line extended out of; for example stimuli, see Fig. 2). As with the other dimensions, the ellipse length dimension also had eight values at equal intervals.

3.1.3. Procedure

The procedure in Experiment 2 was the same as that of Experiment 1, with the exception that the study and test stimuli all consisted of the additional, category-irrelevant dimension, ellipse length. The stimuli were created using the same 64 length-orientation pairings as in Experiment 1, but they were randomly paired with 64 different ellipse length and position pairings.

Participants studied 64 stimuli in a blocked or interleaved schedule, and following this passive study phase, participants were then shown the same 64 length-orientation pairings, but with new ellipse length and position values. Thus, the stimuli were identical to those presented during study with respect to the line lengths and line orientations but were new with respect to the ellipse lengths and positions. Fig. 2 shows an example of the sequencing and stimuli used in Experiment 2. Following each stimulus presentation, participants were asked to select what they believed to be the appropriate category label. This final test was self-paced and without feedback.

3.2. Results and discussion

3.2.1. Classification performance

Average final test performance for each condition is presented in Fig. 4. A 2 × 2 between-subjects ANOVA revealed no main effects, either of category structure, $F(1, 188) = 1.98$, $MSE = 0.03$, $p = 0.16$, or of schedule, $F(1, 188) = 0.01$, $MSE = 0.03$, $p > 0.20$. There was, however, a significant interaction, $F(1, 188) = 7.03$, $MSE = 0.03$, $p = 0.01$, $\eta_p^2 = 0.04$. Post-hoc t-tests revealed that for rule-based categories, accuracy following blocked study ($M = 0.44$, $SD = 0.15$) was significantly higher than accuracy following interleaved study ($M = 0.38$, $SD = 0.13$), $t(81) = 2.12$, $p = 0.04$, $d = 0.43$. The pattern was reversed, however, for information-integration categories. Accuracy following interleaved study ($M = 0.47$, $SD = 0.18$) was marginally higher than accuracy following blocked study ($M = 0.42$, $SD = 0.16$), $t(107) = 1.78$, $p = 0.08$, $d = 0.38$.

3.2.2. Model fits

The model fitting results of Experiment 2 are shown in Table 2. The same decision-bound models from Experiment 1 were applied to these data. Not surprisingly, given the low accuracy rates observed in this experiment, the model fits were poor and the best-fitting model for each participants accounted for a low
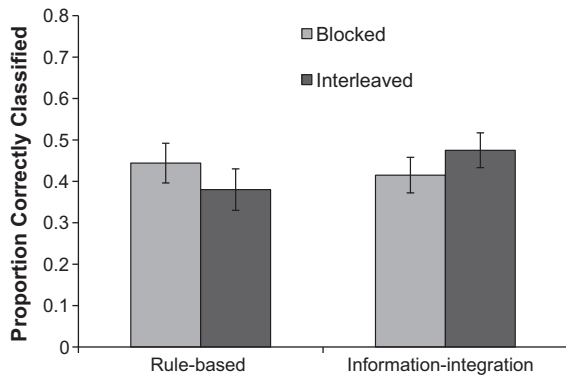
Fig. 4. Classification accuracy results, by schedule and category structure. Error bars represent 95% confidence intervals.

Table 2
Number (percentage) of participants best fit by each type of model for blocked and interleaved study schedules in the rule-based and information-information category structure conditions in Experiment 2.

Best-fit model	Rule-based structure		Information-integration structure	
	Blocked	Interleaved	Blocked	Interleaved
Information-integration	10 (23.2%)	9 (22.5%)	37 (69.8%)	43 (76.8%)
Conjunctive rule	28 (65.1%)	22 (55.0%)	9 (17.0%)	4 (7.1%)
Unidimensional length	2 (4.7%)	3 (7.5%)	2 (3.8%)	3 (0.0%)
Unidimensional orientation	2 (4.7%)	2 (5.0%)	2 (3.8%)	0 (0.0%)
Random responder	1 (2.3%)	4 (10.0%)	3 (5.7%)	6 (10.7%)
Average percentage of responses accounted for by best-fit model	52.3%	45.7%	53.1%	54.9%

percentage of response variability (range: 45.7–54.9%). Even so, the model results converged with those from Experiment 1.

4. Meta-analysis of effects

For rule-based categories, there was a benefit of blocking study of exemplars by category over interleaving study of exemplars from different categories. To estimate the true effect size of the blocking benefit for rule-based categories across, we conducted a meta-analysis[1] of the results from the two experiments, as well as additional data collected as part of a pilot study for a follow-up study.[2] The results of the meta-analysis are presented in the left panel of Fig. 5. The meta-analysis revealed a robust effect of blocking over interleaving for rule-based categories: The estimated effect size (i.e., mean difference, as proportion of correct responses) between the blocked and the interleaved study schedules is -0.07, 95% CI = $[-0.11, -0.03]$, $z(2) = 3.34$, $p < 0.001$. In other words, performance

following blocked study is on average 7% higher than performance following interleaved study, and this advantage of blocking is significantly different from zero. Furthermore, the heterogeneity of the effect sizes was not statistically significant, $Q(2) = 0.23$, $p = 0.89$, $I^2 = 0.0\%$, which indicates that the observed effect did not differ significantly between the three samples.

For the information-integration categories, there was a benefit of interleaving study of exemplars from different categories over blocking study of exemplars by category. To estimate the true effect size of the interleaving benefit for information-integration categories across the two experiments, we again conducted meta-analysis using ESCI. The results are presented in the right panel of Fig. 5. The meta-analysis revealed a robust effect of interleaving over blocking for information-integration categories: The estimated effect size (i.e., mean difference, as proportion of correct responses) between the blocked and the interleaved study schedules is 0.05, 95% CI = $[0.001, 0.10]$, $z(1) = 2.00$, $p = 0.045$. In other words, performance following interleaved study is on average 5% higher than performance following blocked study, and this advantage of interleaving is significantly different from zero. Furthermore, the heterogeneity of the effect sizes was not statistically significant, $Q(1) = 0.10$, $p = 0.92$, $I^2 = 0.0\%$, indicating that the observed effect did not differ significantly between the two studies.

5. General discussion

Across two studies, we tested the effects of blocked and interleaved study schedules on the learning of rule-based and information-integration category structures. Using relatively simple stimuli in Experiment 1 and more complex stimuli in Experiment 2, we found our predicted schedule × category structure interaction: Rule-based category learning benefited from blocking, whereas information-integration category learning benefited from interleaving, and the individual experiment findings are boosted by our meta-analysis results. Modeling results from both experiments suggest that this interaction is mediated by increased rule-based strategy use when category exemplars are blocked.

5.1. What does blocking help participants to learn?

With the rule-based categories, blocking study led to better performance than did interleaving study. In our view, the benefit of blocking for rule-based learning in this particular task may be a result of one of two factors, or both: (1) Blocking may help learners identify the relevant dimensions from the irrelevant dimensions, or (2) blocking may allow learners to generate and test specific hypotheses for each category more easily as they study the exemplars. We conducted a pilot study (first mentioned in footnote 2) as an initial step toward answering this question: In addition to the two conditions from Experiment 2, we compared learning under a third study schedule ($n = 26$) in which the relevant dimensions were interleaved, but the irrelevant ones were blocked (i.e., this schedule was blocked-by-irrelevant-dimensions, as opposed to blocked-by-category), which was designed to draw learners' attention to noticing what dimensions were relevant or irrelevant. On both the classification test and a test in which participants had to identify the relevant and irrelevant dimensions, this new blocked-by-irrelevant-dimensions condition yielded performance at a level comparable to the blocked condition and marginally better compared to the interleaved condition.

Therefore, although we initially hypothesized that participants, when studying one category at a time, are better able to compare exemplars from the same category and to generate and test their hypotheses as to the dimensions define category membership
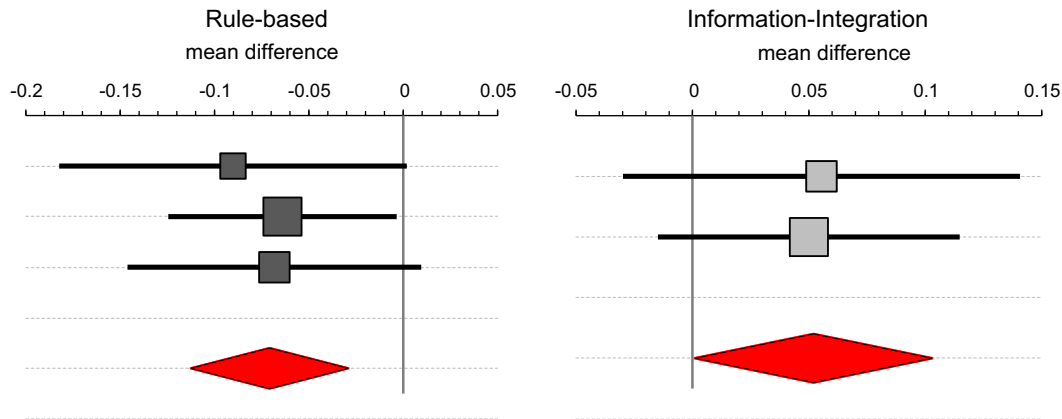
---

[1] The meta-analysis was conducted using the Exploratory Software for Confidence Intervals (ESCI) package (Cumming, 2012, 2014), following the "New Statistics" recommendations by the Association for Psychological Science and Cumming (2012). The package calculates the meta-analyzed effect sizes, by weighting the sample size, group means, and standard deviations across different experiments.

[2] This pilot included the two rule-based category conditions from Experiment 2 ($n = 24$ in each of the schedule conditions), and yielded almost identical results to those of Experiment 2, with blocked condition (M = .45, SD = .14) marginally outperforming the interleaved condition (M = .38, SD = .13), t(46) = 1.77, $p = .08$, $d = .52$. This third replication of a blocking advantage in the rule-based condition gives us even greater confidence in our effect.

Fig. 5. Results of the meta-analyses investigating the effect size of study schedules (mean difference in classification test performance between blocked and interleaved study) for rule-based (left panel) and information-integration (right panel) category structures. The horizontal lines represent 95% confidence intervals for Experiments 1 and 2 (and for the left panel, the third line is the additional pilot data), the location of squares along the x-axis represents the mean differences and the size of the squares represents the weighting of each sample in the meta-analysis. The diamond represents the summary statistic for the mean difference.

(and this may still be true, particularly for Experiment 1), these pilot data suggest that with the addition of irrelevant dimensions (as in Experiment 2), the blocking benefit is perhaps more likely driven by the fact that it allows participants to more easily identify and disregard the irrelevant dimensions. This conclusion, however, is speculative and the present studies do not tease apart these two possibilities.

5.2. What does interleaving help participants to learn?

With the information-integration categories, for which there are no verbalizable rules that optimally distinguish between the categories, interleaving study led to better performance than did blocking study. As information-integration learning does not depend on generating, testing, updating, and maintaining explicit rules, the marginally significant interleaving benefit in information-integration learning may be due to the fact that an interleaved schedule encourages participants to more quickly abandon the use of sub-optimal rule-based strategies during study. When there are a manageable number of to-be-learned categories, it seems plausible that interleaving would enhance learning because it juxtaposes instances of a category with members of other categories, allowing learners to narrow down the defining features of one category that distinguishes it from another. With our design and stimuli, however, we argue that, for rule-based categories, the potential benefit of being able to compare and contrast successive exemplars of different categories that is provided by interleaving is overshadowed by working memory limitations and the costs associated with having to process multiple stimulus dimensions across four categories. The modeling results of Experiment 1 and Experiment 2 provide support for this idea, as interleaved conditions led to both a decrease in rule-based strategy use and increase in information-integration strategy use relative to blocked conditions.

Moreover, the present study is important for another reason: It demonstrates a case where between-category discriminability is not a moderator of the interleaving benefit. The existing literature (e.g., Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Zulkiply & Burt, 2013) on moderators of the interleaving benefit in category learning has largely focused on the discrimination hypothesis: That is, that the interleaving benefit depends on the discriminability of to-be-learned categories and emerges only when between-category discrimination is relatively difficult. Our data suggest, however, that the discriminability hypothesis may not provide a complete account of what determines optimal category-learning

schedules. Since our rule-based and information-integration categories are structurally equivalent, discriminability is equated, yet optimal schedules differ across category structures.

5.3. Possible interactions of discriminability and category structure

It is likely that discriminability and category structure will interact. It has been theorized that the two category learning systems are governed by different factors (e.g., rule-based learning is dependent on working memory, while information-integration learning is not), and thus manipulations of between and within category discriminability could act differently within each system. In other words, it would be overgeneralizing our results to claim that blocking always favors rule-based learning (indeed, we acknowledge that hypothesis testing can proceed from between-category contrasts as well as within-category comparisons, depending on the nature of the to-be-learned, rule-based categories) or that interleaving always favors information-integration learning. Additionally, when there are only one or two to-be-learned categories (for example, category-A and not-A) or when rules are very simple to keep in mind, interleaving might be useful, given the benefits of spacing on memorization (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). When, however, rule-based learning is explicit and thus subject to working memory limitations, as presumably present in our four-dimensional, four-category stimuli set (given overall performance levels), blocked study leads to better learning than does interleaved study.

The discrimination hypothesis and our proposed dual-learning systems framework are not mutually exclusive and future research should explore how these two theories might interact and/or independently contribute to better account for the growing body of literature on sequencing effects in category learning.

6. Concluding comment

Although most real-life categories and concepts cannot be cleanly divided into "rule-based" or "information-integration" categories, the present findings have important implications for education. That is, simply knowing that different types of learning materials may lend themselves more readily to one form of category learning over another may be useful from an educational standpoint. Knowing that a task such as learning artists' styles is less verbalizable and is, therefore, likely to profit more from "information-integration" style learning, for example, whereas learning

to classify organic chemistry compounds, or to classify different types of mathematics and physics problems, is likely to profit more from rule-based learning is useful. Knowing that prior knowledge and expertise of a learner may also play a part, with novices relying more on a rule-based learning system and more advanced learners relying more on an information-integration learning system, is potentially useful as well. Thus, even in our complex and imperfect world, the dual learning-systems framework provides a useful framework for thinking about the methods of instruction that can be used to optimize different types of learning.

Author note

S.M. Noh and V.X. Yan developed the study concept, and developed the study design together with W.T. Maddox. Testing, data collection, and data analysis were performed by S.M. Noh and V. X. Yan, and the data were interpreted under the supervision of W.T. Maddox. S.M. Noh and V.X. Yan drafted the paper, and W.T. Maddox and R.A. Bjork provided critical revisions. All authors approved the final version of the paper for submission.

Acknowledgements

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.
Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale: Erlbaum.
Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 33–53.
Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224*, 147–161.
Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition, 41*, 392–402.
Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition, 42*, 481–495.
Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.
Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York: Routledge.
Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science, 25*(1), 7–29.
Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*, 97–103.
Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585–592.
Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*, 49–70.
Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes, 66*, 309–332.
Maddox, W. T., & Filoteo, J. V. (2005). The neuropsychology of perceptual category learning. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 573–599). Elsevier, Ltd..
Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 227–235.
Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition, 108*, 578–589.
Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews, 32*, 279–291.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.
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.
Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41*, 16–27.