

Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study

Paulo F. Carvalho · Robert L. Goldstone

Published online: 3 October 2013
© Psychonomic Society, Inc. 2013

Abstract Recent research in inductive category learning has demonstrated that interleaved study of category exemplars results in better performance than does studying each category in separate blocks. However, the questions of how the category structure influences this advantage and how simultaneous presentation interacts with the advantage are open issues. In this article, we present three experiments. The first experiment indicates that the advantage of interleaved over blocked study is modulated by the structure of the categories being studied. More specifically, interleaved study results in better generalization for categories with high within- and between-category similarity, whereas blocked presentation results in better generalization for categories with low within- and between-category similarity. In Experiment 2, we present evidence that when presented simultaneously, between-category comparisons (interleaved presentation) result in a performance advantage for high-similarity categories, but no differences were found for low-similarity categories. In Experiment 3, we directly compared simultaneous and successive presentation of low-similarity categories. We again found an overall benefit for blocked study with these categories. Overall, these results are consistent with the proposal that interleaving emphasizes differences between categories, whereas blocking emphasizes the discovery of commonalities among objects within the same category.

Keywords Interleaving · Inductive learning · Perceptual category learning · Comparison

Electronic supplementary material The online version of this article (doi:10.3758/s13421-013-0371-0) contains supplementary material, which is available to authorized users.

P. F. Carvalho (✉) · R. L. Goldstone
Department of Psychological and Brain Sciences, Indiana University,
1101 E. 10th Street, Bloomington, IN 47405, USA
e-mail: pcarvalh@indiana.edu

How to present information so that learning and memory are optimized is an important issue in teaching and training contexts (Rohrer & Pashler, 2010). It has long been demonstrated that spacing repeated presentations of the same information results in better memory than does repeating the same information several times within a single occasion, even when time and number of presentations are equated (Ebbinghaus, 1913). This memory phenomenon, known as the “spacing effect,” is a highly robust finding (Delaney, Verkoeijen, & Spiguel, 2010; Proctor, 1980) that has been shown both in experimental situations with words and pictures and in more applied situations, such as flashcard studying (Kornell, 2009).

This previous work on spacing presentations is applicable to learning contexts involving the storage and retrieval of factual information. However, the importance of maximizing memory for specific facts will often not be as educationally relevant as learning general concepts with an open-ended number of items. Whereas rote memorization can be used to learn the factual knowledge that veins lead the blood to the heart, open-ended concept learning and induction is required to reliably categorize cross-sectional slides of arteries and veins. Pedagogically speaking, it is often important to know whether the way that instances are presented influences inductive learning and subsequent generalization of the acquired knowledge.

The question of how to present information in order to optimize category learning and generalization has been raised before, and several proposals have been put forward. Some of these proposals are related to the categories being taught. For example, Elio and Anderson (1984; see also Sandhofer & Dumas 2008) have proposed that learning should start with low-variability items and that items with greater variability should be introduced later. Another proposal is that items that present the same generalization should be presented close together in temporal sequence (Elio & Anderson, 1981; Mathy & Feldman, 2009). Other proposals are related to categorization

difficulty. One such proposal is that items that previous learners had difficulty categorizing should be presented early to new learners (Lee, MacGregor, Bavelas, & Mirlin, 1988) or that study should be sequenced in increasing order of complexity, from simple to complex examples (Hull, 1920; but see Spiering & Ashby, 2008).

Interleaved versus blocked study

Researchers have also proposed interleaving items from the categories being taught—that is, presenting one item from one category followed by an item from another category, rather than grouping items from the same category together (blocking). Recent results have shown that alternating presentations of the categories leads to better inductive learning and memory than when the categories are presented in separate blocks.

This advantage of alternating to-be-learned categories has been observed for different kinds of concepts, such as artists' styles (Kang & Pashler, 2012; Vlach, Sandhofer, & Kornell, 2008), bird species (Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011), novel category learning in children (Vlach et al., 2008) and older adults (Kornell et al., 2010; Wahlheim et al., 2011), and mathematical operations in primary school students (Taylor & Rohrer, 2010).

Initial accounts of this phenomenon related the advantage of interleaving categories with the greater temporal spacing introduced between repetitions of the same category (Kornell et al., 2010). However, recently Kang and Pashler (2012; see also Birnbaum, Kornell, Bjork, & Bjork, 2013) presented participants with several paintings from different artists and used a procedure similar to those from previous studies (Kornell & Bjork, 2008), but with added presentation conditions. In one experiment, the authors compared categorization performance in a generalization test preceded by one of four study conditions: (1) blocked, (2) interleaved, (3) blocking in which every presentation of a painting was followed by an unrelated filler task (temporal spaced condition), and (4) simultaneous presentation of all exemplars from the same artist (blocked simultaneous condition). The results showed that only the interleaved condition resulted in better performance than did the blocked condition, thus providing evidence that greater temporal spacing of presentations is not the critical factor for the interleaved advantage (but see Birnbaum et al., 2013).

An alternative explanation is that interleaving maximizes inductive learning by promoting discrimination between stimuli of the different categories (Goldstone, 1996; Kang & Pashler, 2012; Kornell & Bjork, 2008; Mitchell, Nash, & Hall, 2008). In the blocked condition, the juxtaposition of exemplars of the same category would not yield any advantage because

finding differences between those stimuli is not beneficial for learning.

However, in some situations interleaving is not ideal, and previous studies have also demonstrated an advantage of blocked presentation. For example Goldstone (1996) presented participants with complex images composed of 20 line segments. Eight of the segments were diagnostic (tending to be present in only in one of the categories), whereas the other 12 were nondiagnostic (tending to be present in both categories). The study included two conditions: frequent alternation of categories (interleaving¹) and infrequent alternation (blocking). Participants had to classify each stimulus in one of the two categories with corrective feedback. The results showed that participants were better at learning the categories in the infrequent-alternation condition. The author associates this advantage with the relative difficulty in finding the common features shared by the members of each category (for similar results with different stimuli, see Kurtz & Hovland, 1956; Whitman & Garner, 1963).

Given this mixed evidence of what is the best way to present information for optimal learning, one potentially important question is, What conditions yield an advantage for interleaving as compared to blocking? In the present work, we tried to provide an initial answer to this question by manipulating the properties of the stimuli being learned and the temporal dynamics of the task.

Goldstone (1996) proposed that category learning might be difficult for two different reasons: high between-category similarity or low within-category similarity. *High between-category similarity* refers to category structures in which the different categories share most of their features, making discriminating the categories a matter of finding subtle differences between exemplars (as is the case for distinguishing between, e.g., alligators and crocodiles). *Low within-category similarity*, on the other hand, refers to category structures in which the exemplars within one category share very few features (as is the case for, e.g., the category “animal”).

Each one of these two kinds of category structures requires different processes for efficient category learning. In the case of high between-category similarity, the central challenge is to identify subtle differences between categories, which might be facilitated by frequent alternation among the categories. For example, distinguishing alligators from crocodiles requires attention to relatively subtle features that discriminate between

¹ This manipulation of blocking/interleaving differs from the previously described experiments. An advantage of alternating successively presented categories either frequently (75%) or rarely (25%) is that the category-level predictability of the two conditions is equated, while still establishing a challenging category-learning task. A disadvantage is that both blocked and interleaved conditions will have category members that are temporally distributed over the course of training. Given that our principal goal in the present work was to investigate the role of sequential comparison in category learning and generalization, this disadvantage was not deemed overly serious, and we used this approach.

the species, and alternating between these species facilitates this. However, rapid alternation between two categories with low within-category similarity may not facilitate the identification of relevant properties of each category. In this case, it might be more beneficial to block exemplars separately by each category, so that the learner can identify the shared features among members of a category hidden within their diversity (Goldstone, 1996). For example, to categorize a physics problem as requiring classical versus quantum theoretical constructs, it may help to train students first on a block of one type of problem and then to present the other type of problem. By blocking training, the subtle theoretical assumptions and constructs of each type of problem may be more clearly highlighted by comparing successive problems that share these features (see Homa & Chambliss, 1975, for similar proposals relating to number of categories and size of each category and its effects on highlighting common and discriminating features).

Simultaneous presentations and interleaved versus blocked study

Another important question is related to the temporal factors involved in interleaving versus blocking. Even though the interleaved advantage has been shown to be more dependent upon the juxtaposition of different categories than increased spacing (Kang & Pashler, 2012), previous research has demonstrated that category learning can be facilitated by changing the presentation delay between different categories or exemplars of the same category—for example by presenting instances simultaneously. The role of simultaneous presentation in category learning has been emphasized before in research with children (Gentner & Namy, 1999; Graham, Namy, Gentner, & Meagher, 2010; Kotovsky & Gentner, 1996; Namy & Gentner, 2002; Sims & Colunga, 2010; Vlach, Ankowski, & Sandhofer, 2012), and adults (Hammer, Bar-Hillel, Hertz, Weinshall, & Hochstein, 2008; Higgins & Ross, 2011; Spalding & Ross, 1994).

The main finding is that when two or more exemplars of a category are presented simultaneously, participants are better at identifying the relevant features for categorization, even if those features are less salient than irrelevant ones. Additionally, the advantage for simultaneous presentation can also be seen in discrimination learning. When learning to discriminate two similar objects, presenting both simultaneously results in better discrimination than presenting them successively, either interleaved or blocked (Dwyer, Mundy, & Honey, 2011; Mundy et al., 2007, 2008). One possible reason for this advantage is the reduced memory constraints as compared to successive presentation. This would allow learners to more effectively compare the two objects and extract the relevant information (Andrews, Livingston, & Kurtz, 2010).

Traditionally the study of interleaving/blocking schedules of presentation has been done with successive presentations. Another alternative, pursued here, is to have learners categorize object one at a time, but to simultaneously display the previously categorized object with its correct category assignment as well as the current object to be categorized. This manipulation reduces the time delay between interleaved presentations without increasing the amount of information available to participants (participants continue having access only to two of the category exemplars at each moment). It is a promising method for concept training because its categorization-with-feedback protocol encourages an active problem-solving attitude for learners, whereas the simultaneity of previous and current objects may facilitate comparison between them.

How does reducing the temporal delay between successive categorization decisions during category learning influence which presentation schedule results in better learning for different kinds of categories? One possibility is that simultaneous presentation of a previously categorized stimulus along with a new one will reduce the memory load associated with remembering the characteristics of the previous item, maximizing the benefits of interleaving for high-similarity categories and blocking for low-similarity categories. Another possibility is that this simultaneous presentation will introduce new constraints. For instance, simultaneous presentation may emphasize differences between objects that belong to different categories (Lipsitt, 1961). This could change how participants solve the learning task, thus changing how interleaved or blocked study affect learning.

The present work

In this research, we investigated whether the advantage of interleaved over blocked study depends on the characteristics of the learning situation such as the categories being learned and the time delay between successive presentations. In Experiment 1, we showed that interleaved study of three categories results in better generalization for high-similarity categories but blocked study results in better generalization for low-similarity categories. In Experiment 2, we showed that when learners simultaneously view the preceding categorized item along with a novel one, interleaving categories results in better generalization for high-similarity categories, but there is no difference between the two schedules for low-similarity categories. Experiment 3 further explored the effect of simultaneity for low-similarity categories and replicated the results of Experiments 1 and 2.

Experiment 1

Previous demonstrations of the advantage of interleaved study have used highly similar categories (e.g., Kang & Pashler,

2012; Kornell & Bjork, 2008; Kornell et al., 2010; Wahlheim et al., 2011). In this experiment, we directly contrasted the effects of interleaved over blocked study in categories with different similarity structures. The objective of this manipulation was to reconcile the apparently contradictory evidence that both blocked and interleaved study can be beneficial (see the introduction).

Two sets of categories were used: a high-similarity set, in which exemplars had both high within- and between-category similarity, and a low-similarity set, in which exemplars had both low within- and between-category similarity. Participants studied categories from one of these sets in an interleaved and blocked fashion. Following Goldstone's (1996) proposal, we predicted that interleaved study would result in better generalization performance for high-similarity categories. Conversely, for low-similarity categories, generalization performance was expected to benefit from blocked study of the categories.

Method

Participants

A group of 76 Indiana University undergraduate students participated in this experiment in return for partial course credit. All participants completed both the blocked and interleaved conditions, using different categories. Of the participants, 44 completed the high-similarity condition and 32 the low-similarity condition. Fifteen participants (all in the high-similarity condition) did not reach the criterion of 34% or more correct responses during categorization learning in one or both of the schedule conditions and were excluded from further analyses.²

Stimuli and apparatus

The stimuli used in this and the following experiments were blob figures (see Fig. 1). All blobs were created by randomly generating curvilinear segments. A single curvilinear segment defined each category and was present in all exemplars of that category. Across all of our experiments, two sets of six categories were used, a low-similarity set and a high-similarity set, for a total of 12 categories. Each category was composed of 16 exemplars.

In the high-similarity set, exemplars shared most of their features with all the other exemplars in the same category and in each of the other five categories. Moreover, variation within each category was exactly the same for all categories, so that a

² This is a relatively high proportion of participants who failed to achieve an overall performance above the chance level of 34%. This reflects the difficulty level of learning to discriminate between high-similarity categories.

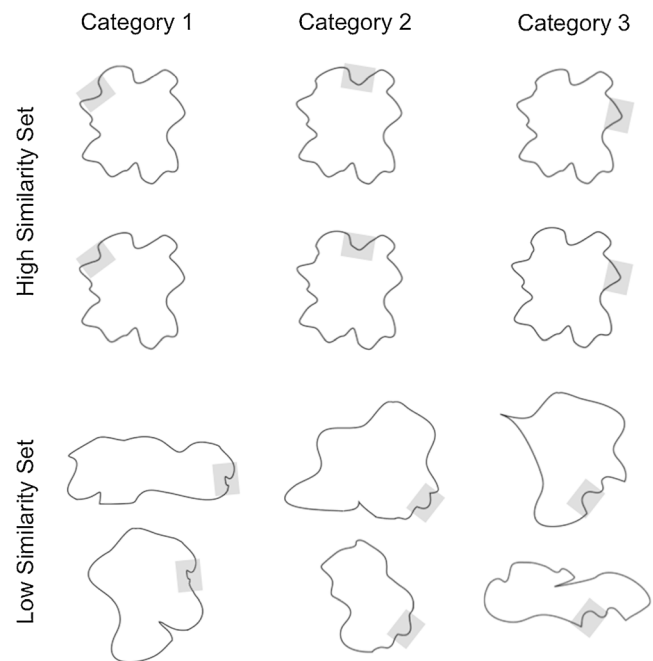


Fig. 1 Sample stimuli from three categories in each stimulus set. The top six stimuli are from the high-similarity set, and the bottom ones are from the low-similarity set. A shaded area indicates the diagnostic feature for each category. This shaded area is for illustration purposes only and was not presented to participants

difference that could exist between two exemplars in Category 1 would also exist between two exemplars of each of the other categories in the set.

In the low-similarity set, exemplars within each category shared only the category-relevant feature.³ Moreover, exemplars from different categories differed in all their features. Some of the exemplars had an overall round shape, and others an overall oblique shape (this variability was equally distributed across categories).

As a cover story, participants were told that a recent expedition to Mars had recovered several cells of alien organisms. Each cell could be categorized into one of three species solely on the basis of its perceptual features. Stimuli were presented on a computer screen, and participants responded using keys on the keyboard with a consistent mapping to the category assignment.

Procedure

In this experiment, we manipulated the similarity structure of the categories being studied (low vs. high similarity,

³ Of course, all of the objects do have some features in common, such as being closed shapes, outlined in black, curvy, and white in the middle. Similarly, what we experimentally manipulated as one feature may have been construed as multiple features by the participants, because each feature involved multiple curves. We cannot be sure about the exact number of psychological features contained in our stimuli. The logic of our experiment only depended on the relative number of shared features decreasing from the high- to the low-similarity sets.

manipulated between subjects) and the schedule of presentation during study (interleaved vs. blocked presentations, manipulated within subjects). Each participant was assigned to one of the similarity conditions, and all completed both schedule conditions. Each schedule condition was composed of two identical phases: a study task and a generalization task (always presented in this order).

Study task In this task, participants were presented with stimuli from each of three categories. Participants were presented with a stimulus in the center of the screen for 500 ms. After the blob was removed, the participant was asked to classify the blob into one of three species (Q, Y, and P, or A, G, and L) by pressing the corresponding key on the keyboard. After the participant's response, the blob was presented again for 2,000 ms, together with the presentation of corrective feedback (e.g., "CORRECT! This cell belongs to species Q" or "Sorry, that is INCORRECT! This cell belongs to species Y"). A 1,000-ms intertrial interval followed and then a new trial began.

In the blocked condition, the categories presented alternated 25% of the time, whereas in the interleaved condition, they alternated 75% of the time. Thus, in the interleaved condition, the probability of a blob being followed by a blob of the same category was low, whereas for the blocked condition, this probability was high. We used this probabilistic approach rather than creating purely interleaved or blocked conditions in order to diminish the possibility that participants noticed the pattern of alternation in responses, which would affect categorization accuracy. Furthermore, if a purely blocked condition had been used, there would be no way to guarantee participants' attention to the task, as they would have no uncertainty as to the correct categorization. This approach has been used before in similar tasks, with successful results (Goldstone, 1996).

The two study conditions (blocked vs. interleaved) differed only in the frequency of category change and the species labels. In one of the conditions, Q, Y, and P labels/keys were used, whereas in the other, A, G, and L were used, by random assignment. Which condition was presented first was counterbalanced across participants and the allocation of the stimuli to each category and condition was randomized across participants.

Participants completed four blocks of this task, each one composed of 72 trials (three presentations of each of the eight stimuli from each of the three categories).

Generalization task This second phase was a generalization task during which 48 stimuli were shown in random order—the 24 blobs participants saw during the study task and 24 new stimuli. The new stimuli were generated in the same manner as the training stimuli, with new instantiations of the unique features. Each stimulus was presented in the center of the screen for 500 ms, after which participants were asked to

classify it into one of the species just learned. After a 1,000-ms intertrial interval, a new trial would begin. No feedback was provided during this phase.

Results and discussion

In these and all subsequent analyses, order effects (whether participants started with the blocked or interleaved condition) were analyzed and no effect of order of conditions was found.

First, we report performance over the study phase, as is depicted in Fig. 2. As can be seen from the graphs, for both low- and high-similarity category structures, performance increases across blocks and is superior for the blocked condition as compared to the interleaved condition. Moreover, performance for the low-similarity structure is overall superior to that of the high-similarity structure.

A mixed analysis of variance (ANOVA) with Condition and Block as within-subjects factors and Similarity Structure as a between-subjects factor confirms this interpretation, revealing a main effect of similarity structure, $F(1, 58) = 37.64$, $MSE = .091$, $p < .0001$, and schedule condition $F(1, 58) = 45.44$, $MSE = .041$, $p < .0001$, but no interaction between the two, $F(1, 58) < 1$, $MSE = .041$. The main effect of block was also significant, indicating that for both similarity structures and presentation conditions, participants improved their categorization, $F(3, 174) = 224.30$, $MSE = .0058$, $p < .0001$. Moreover, this improvement was more marked for the low- than for the high-similarity structure, $F(3, 174) = 18.77$, $MSE = .0058$, $p < .0001$, and the improvement was greater for interleaved than for blocked study, across both similarity structures, $F(3, 174) = 6.35$, $MSE = .0053$, $p = .0004$.

Thus, from the analyses of performance during the study task, we would conclude that blocking results in better performance overall for both category structures, and that low-similarity categories are easier to learn overall. From these conclusions we could predict any of three possible outcomes for generalization performance: (1) Interleaving constitutes a more demanding study format, implicating greater cognitive effort and thus constituting a "desirable difficulty" (Bjork, 1994). This interpretation would predict better generalization performance in the interleaved condition regardless of category structure; (2) blocking results in overall better learning and will result in equally better performance during generalization (Kurtz & Hovland, 1956; Whitman & Garner, 1963); and (3) even though blocking results in better performance during study, this performance does not necessarily imply performance at generalization. By this interpretation, the successful acquisition and encoding of the categories is likely to depend on several factors including the category structure and the way items are presented during study.

The results from the generalization task are depicted in Fig. 3. As can be seen from the graph, no overall main effect of schedule of presentation is apparent, and performance is

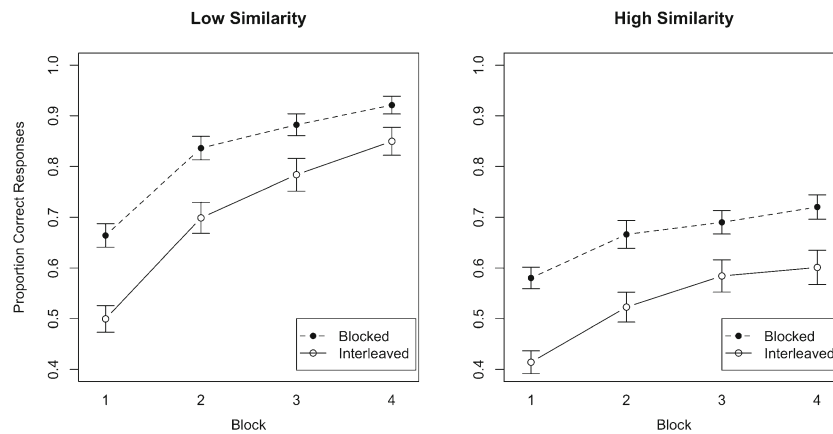


Fig. 2 Results for the study phase of Experiment 1. Error bars indicate standard errors of the means. Chance-level performance in this task was .33

higher overall for low-similarity categories. Moreover, and of greater interest, for novel stimuli, blocked presentation during study improves performance for low-similarity categories. Conversely, for high-similarity categories, interleaving study results in better generalization performance to novel items.

A mixed ANOVA with Category Structure as a between-subjects factor and Schedule of Presentation and Type of Stimuli (novel vs. studied) as within-subjects factors confirms this interpretation. Performance is overall higher for low-similarity categories, $F(1, 58) = 25.08$, $MSE = .094$, $p < .0001$, and across category structures performance is overall higher for studied stimuli, $F(1, 58) = 62.56$, $MSE = .015$, $p < .0001$. Interestingly, we observed no overall advantage of one schedule over the other, $F(1, 58) = 1.53$, $MSE = .048$, $p = .217$. However, an interaction did emerge between category structure and schedule of presentation, $F(1, 58) = 5.99$, $MSE = .048$, $p = .017$, indicating that for high-similarity categories interleaving is overall better, but for low-similarity categories there is no overall advantage of one schedule over the other. The interaction between the schedule of presentation, the category structure and the type of item is also significant, $F(1, 58) = 8.32$, $MSE = .0097$, $p = .005$. To further analyze this critical interaction, we calculated the difference in performance between the two study schedules for each type of item by subtracting performance for the blocked condition from performance for the interleaved

condition for each participant. This analysis is plotted in Fig. 4: Negative values indicate a benefit for blocked study, whereas positive values indicate a benefit for interleaving study.

As the plot in Fig. 4 shows, for high-similarity categories, interleaved study results in better generalization performance to novel stimuli, but for low-similarity categories, blocked study results in better generalization performance, $t(58) = 3.00$, $p = .004$, $d = 0.76$. However, for studied items we found no effect of category structure, $t(58) = 1.26$, $p = .213$, $d = 0.32$.

The results of this experiment show that (a) performance during the study phase as a function of schedule is not a reliable predictor of later generalization performance and (b) the benefit of interleaved over blocked study is dependent on contextual factors such as the category structure.

It has been demonstrated before that higher performance during study does not necessarily result in better learning (Taylor & Rohrer, 2010). One general explanation for dissociated study–test performance is that the greater variability associated with interleaving items leads to a more difficult categorization and implicates greater cognitive resources, thus constituting a desirable difficulty (Bjork, 1994). By this account, interleaving should always result in better performance, which is not the case here.

In fact, blocking resulted in overall higher performance during study for both high- and low-similarity category

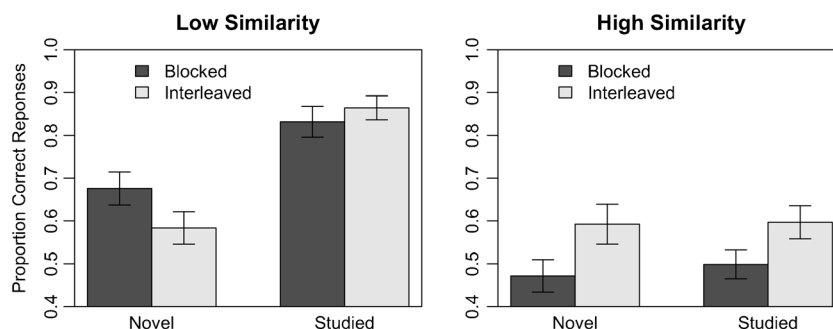


Fig. 3 Results for the generalization phase of Experiment 1. Error bars indicate standard errors of the means. Chance-level performance in this task was .33

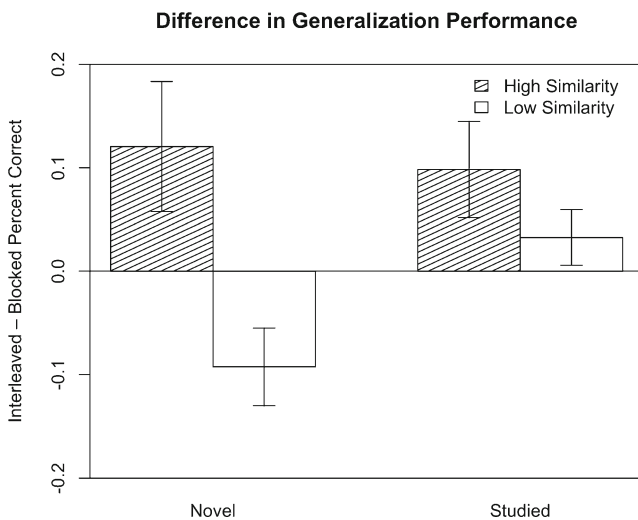


Fig. 4 Differences in generalization performance between the interleaved and blocked conditions. The differences were calculated for each participant by subtracting performance following blocked study from performance following interleaved study. Higher values indicate a benefit of interleaving over blocking. *Error bars* represent standard errors of the means

structures, which might indicate the use of a heuristic such as “Use the previous object’s categorization for the present object” in the blocked condition. The use of this heuristic would allow participants to achieve 75% accuracy, whereas the equivalent heuristic of “Guess a different category than the previous object’s categorization” in the interleaved condition would only allow for 37.5% accuracy (the study included three categories and 25% repetition, $.5 * .75 + 0 * .25$). This might indicate that blocked study results in worse learning because participants are using these heuristics instead of studying the stimuli. However, even though the use of these heuristics was possible, and even likely, in both the high- and low-similarity category sets, only in the high-similarity condition did we observe an interleaved advantage for novel item generalization. In fact, for low-similarity categories the opposite occurred: Blocked study resulted in improved generalization to novel items. This is contrary to the sole use of a heuristic without attending to the stimuli and the use of such heuristics cannot explain the most theoretically important results presented here. It is possible that some of the advantage for interleaved study with high-similarity categories is the result of decreased processing and a more extensive use of heuristics in the blocked condition. However, we would argue that the use of these heuristics is related to the difficulty in identifying the discriminating features necessary to obtain good performance. In this sense, learners might rely on other means, such as maintaining the same answer, in order to achieve good performance. The fact that we found an advantage for blocked study with low-similarity stimuli is in agreement with this explanation. Another reason to believe it is

unlikely that participants in the blocked condition are relying only on these heuristics is the fact that previous research, using different methodologies that do not allow for the use of these heuristics, found similar results when high-similarity categories were used (e.g., Kornell & Bjork, 2008; Wahlheim et al., 2011). Additionally, notice that it is unlikely that participants were using an exemplar-based categorization approach in this task, given that the advantage was mostly seen for novel stimuli but not for studied ones.

Mitchell, Nash, and Hall (2008) propose an account of interleaved study increasing the salience of the category discriminating features relative to the other features. This saliency increase takes place because the discriminating feature is not repeated on every trial, whereas all of the other features are. Although this can indeed account for the results seen for the high-similarity set, it does not easily account for the results seen for the low-similarity set. In fact, in the low-similarity set the exact opposite seems to take place: Most features change from trial to trial, and participants benefit from blocking—that is, the repetition of the category-discriminating feature.

One possible explanation for these results is that the two study schedules emphasize similarities and differences among successive stimuli differently. In this way, which of the two schedules is more beneficial will change depending on whether learning the category requires participants to find similarities among objects of the same category (as for categories in the low-similarity set) or find differences between objects of different categories (categories in the high-similarity set).

Experiment 2

Experiment 1 demonstrated that whether interleaved or blocked study is more advantageous for learning is a function of the similarity structure of the categories being studied. In this experiment, we altered the dynamics of the study task by simultaneously presenting the previously categorized object along with a novel, to-be-categorized one.

This manipulation reduced the memory constraints of the task, but might also introduce new constraints related to how participants go about solving the study task. If the effects seen in Experiment 1 were due to comparison of successive stimuli, we could amplify the advantage of each study schedule by promoting comparison through simultaneity. However, it is also possible that an overall amplification will not occur. It is possible that some comparisons are selectively emphasized through simultaneous comparison—for example, features that differ between objects belonging to different categories (see, e.g., Lipsitt, 1961; MacCaslin, 1954). In this way, we sought to investigate in which ways having direct access to the previous stimulus affected the advantage of each study schedule for low- and high-similarity categories.

Method

Participants

A group of 118 Indiana University undergraduate students participated in this experiment for partial course credit. All participants completed both the blocked and interleaved conditions, using different categories. Of these participants, 61 completed the high-similarity condition, and 57 the low-similarity condition. Twenty-three participants (16 in the high-similarity condition and seven in the low-similarity condition) did not reach the criterion of 34% or more correct responses during categorization learning in one or both of the schedule conditions and were excluded from further analyses.

Stimuli and procedure

This experiment followed a procedure similar to that of Experiment 1, except for the following changes (see Fig. 5 for screenshots of the study phase for Exp. 2). During the study phase, participants saw a stimulus on the right side of the screen that they had to classify as belonging to one of the three species of alien cells. After the participant's response, the

stimulus remained on the right side of the screen for another 2,000 ms, with feedback and correct category assignment above it. Next, the stimulus and correct category assignment moved via an animation to the left side of the screen and remained there for the duration of the subsequent trial. In this way, on any categorization trial (excluding the first one), the participant could see the previous stimulus and category feedback as well as the stimulus to be categorized, simultaneously on the screen.

Additionally, before the beginning of the generalization task, participants performed a practice phase of four trials similar to the trials in the generalization task but presenting pictures of real objects. This additional practice phase was introduced in this experiment because initial pilots revealed that the transition from the self-paced study phase to the fast-paced generalization task resulted in the loss of the initial trials due to participants' adjustment to the new type of task.

Results and discussion

We started by analyzing the data from the study phase (see Fig. 6). As can be seen from the graph, we observed an increase in accuracy across blocks, and this learning is more accelerated

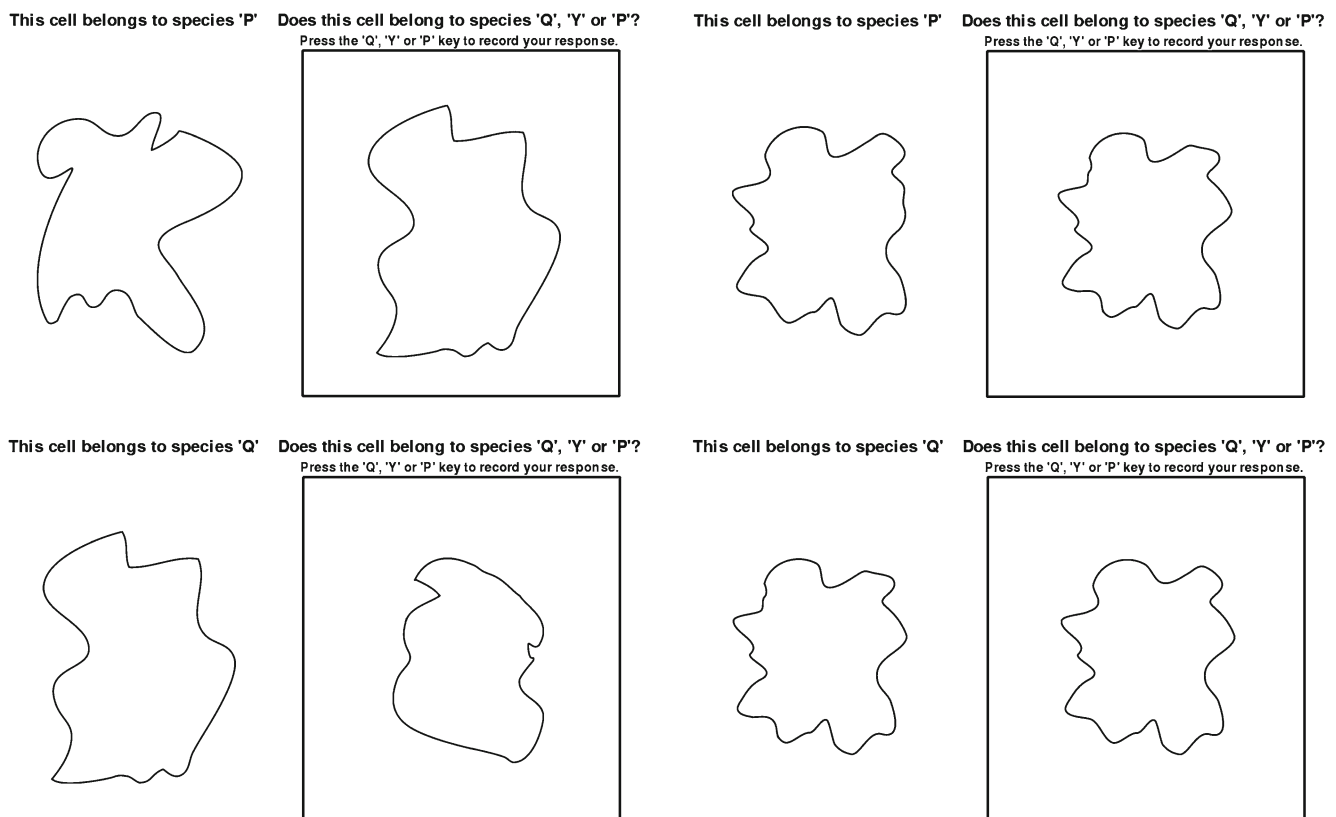


Fig. 5 Screenshots of trials t (bottom panel) and $t - 1$ (top panel) in the study phase of Experiment 2 (the low-similarity condition on the left, and the high-similarity condition on the right). Participants had to classify the stimulus on the right by pressing the corresponding key. After the participants'

responses, feedback was presented above the stimulus on the right, and then the stimulus moved to the left along with the correct assignment. The stimulus on the left in the bottom panel screenshot (trial t) is the one presented in the previous trial (trial $t - 1$)

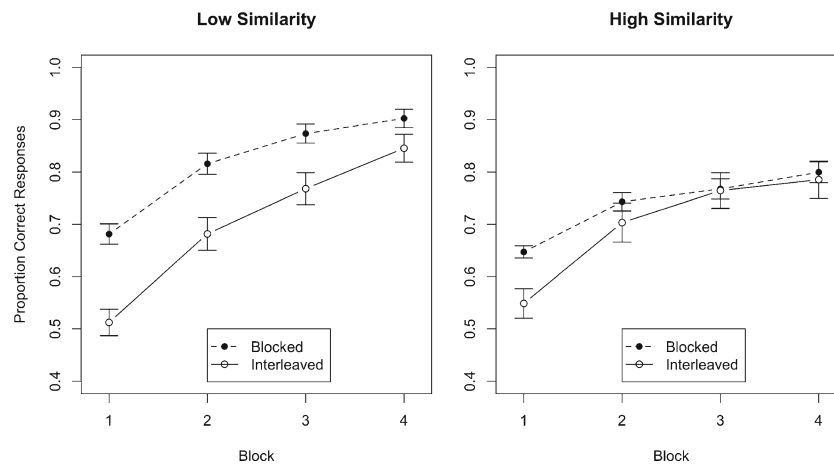


Fig. 6 Results for the study phase of Experiment 2. Error bars indicate standard errors of the means. Chance-level performance in this task was .33

for low-similarity than for high-similarity categories. Moreover, there is an overall advantage of blocking in performance, but this advantage seems to be lost by the last two blocks of study for high-similarity categories. No overall difference in performance emerged between the low- and high-similarity conditions.

A mixed ANOVA with Block and Presentation Schedule as within-subjects factors and Category Structure as a between-subjects factor confirmed this interpretation. We found main effects of schedule, $F(1, 92) = 20.27$, $MSE = .059$, $p < .0001$, and block, $F(3, 276) = 276.55$, $MSE = .0074$, $p < .0001$, but no main effect of category structure, $F(1, 92) = 2.09$, $MSE = .14$, $p = .152$. However, we did see significant interactions between category structure and schedule, $F(1, 92) = 4.72$, $MSE = .059$, $p = .032$, and between category structure and block, $F(3, 276) = 8.11$, $MSE = .0074$, $p < .0001$. Finally, a significant interaction emerged between schedule of presentation and block, $F(3, 276) = 13.80$, $MSE = .0064$, $p < .0001$, indicating that the improvement in performance over the study phase was greater for the interleaved condition.

This pattern is similar to the one seen in Experiment 1. As we saw in Experiment 1, overall blocking seems to result in improved performance. However, simultaneous presentation reduced this advantage for high-similarity categories. Of greater interest is what effect simultaneous study had on generalization for low and high-similarity categories.

The results from the generalization task are shown in Fig. 7. As in Experiment 1, generalization seems to be better overall for low-similarity categories. Moreover, an interaction is apparent between category structure and schedule of presentation, with interleaved study resulting in better performance for high-similarity categories but for low-similarity categories there are no differences in performance between the two schedules. This is surprising, given the results of Experiment 1 and the large difference between blocking and interleaving for the low-similarity category structure seen during study.

A mixed ANOVA with Stimulus Type and Presentation Schedule as within-subjects factors and Category Structure

as a between-subjects factor confirmed the main effect of category structure, $F(1, 92) = 4.02$, $MSE = .14$, $p = .048$. We also observed a main effect of schedule of presentation, $F(1, 92) = 10.80$, $MSE = .053$, $p = .001$, with interleaving resulting in overall better performance than blocking. Performance was also higher for studied stimuli than novel ones, $F(1, 92) = 43.62$, $MSE = .012$, $p < .0001$. This analysis also confirmed our initial interpretation that the schedule of presentation seems to have an effect only for high-similarity categories, $F(1, 92) = 14.55$, $MSE = .053$, $p = .0002$.

The results of this experiment show that, when learners have the opportunity to simultaneously study the to-be-categorized item and the previously categorized item, the interaction between category structure and presentation schedule changes. In this situation, the advantage for blocking over interleaving for low-similarity categories is lost, and the advantage for interleaving for high-similarity categories is preserved and even numerically greater. However, we still observed an interaction between category structure and presentation schedule: For high-similarity categories, interleaved study resulted in higher performance, whereas for low-similarity categories, the way that exemplars were studied did not seem to affect learning.

One possible reason for these results is that the simultaneous presentation design presented here increased the number of participants who relied solely on the heuristic “choose the same category as in the previous trial” during the study phase. Evidently, participants could also use this heuristic in Experiment 1, and the overall better performance for the blocked condition across the two experiments seems to indicate that they did, to some degree. However, the presence of the object from the previous trial along with its correct assignment on the screen is likely to increase the use of this heuristic. Furthermore, because there is no ambiguity (due to memory failures) about what the previous category was, the effectiveness of the heuristic might even be higher in Experiment 2 than in Experiment 1. The increased reliance on

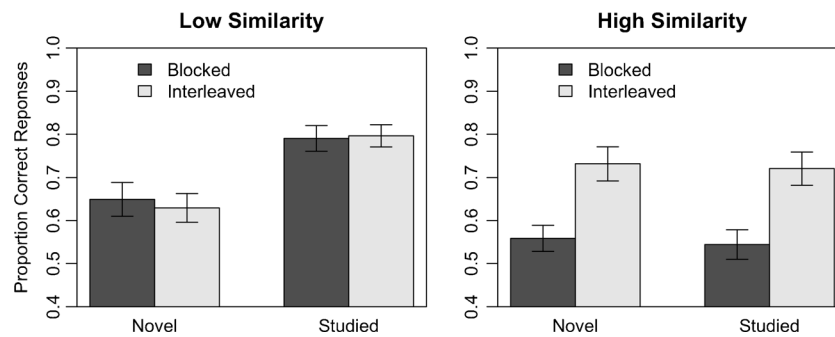


Fig. 7 Results for the generalization phase of Experiment 2. Error bars indicate standard errors of the means. Chance-level performance in this task was .33

this heuristic would result in greatly decreased learning for the blocked condition because participants are not actively trying to identify the categories' properties, but rather simply using on-screen information about the previous item's category.

We investigated this hypothesis by calculating the proportions of times that participants chose the same category on trial t and $t - 1$, when the category actually changed (response stay), and the proportions of times that participants chose a different category on trial t relative to trial $t - 1$, when the two stimuli were in fact from the same category (response change). If participants were aware of the transition probabilities, we expected the proportions of response stays to be higher for the blocked condition and the proportions of response changes to be higher for the interleaved condition (i.e., being somewhat biased to answer "same" or "different," depending on the condition). Moreover, we looked at differences between Experiments 1 and 2 in this measure of bias. These analyses revealed that participants internalized the probabilities of each schedule, but this did not seem to change between the successive and simultaneous experiments [in the blocked condition, $M_{\text{Stay}} = .22$ ($SD = .27$) for Exp. 1 and $M_{\text{Stay}} = .20$ ($SD = .27$) for Exp. 2, and in the interleaved condition, $M_{\text{Change}} = .23$ ($SD = .27$) for Exp. 1 and $M_{\text{Change}} = .19$ ($SD = .29$) for Exp. 2, both p s > .05].

Another possibility is that simultaneous presentation increases the saliency of the differences between stimuli from different categories, improving performance following interleaved study. In the **General Discussion**, we examine this possibility to account for the different results between the successive and simultaneous experiments.

However, given the differences found between Experiments 1 and 2 regarding the benefit of blocked study for generalization of low-similarity categories, it was important to directly compare simultaneous and successive presentations of low-similarity categories. For this purpose we conducted Experiment 3.

Experiment 3

In this experiment, we directly compared simultaneous and successive study of low-similarity categories. The main

purpose of this study was to investigate the effect of simultaneity on low-similarity categories. For this purpose, we developed an experiment in which only low-similarity categories were studied, both blocked or interleaved (manipulated within subjects), and exemplars could be studied simultaneously or successively (manipulated between subjects). This allowed us to directly compare simultaneous and successive presentations in the study of low-similarity categories.

Additionally, one alternative explanation for the results presented thus far is that participants in both experiments may possibly have learned the pattern of key presses during study and relied solely on these. Although this hypothesis would not explain the interaction found between schedule of presentation and category structure in either Experiment 1 or 2, in the present experiment we changed the way categories were studied. In this experiment, participants studied the exemplars along with the correct category assignment.

Overall, we expected to find a benefit for blocked study in the successive group as seen in Experiment 1 and not a benefit for either schedule for the simultaneous group (as in Exp. 2).

Method

Participants

A total of 96 Indiana University undergraduate students participated in this experiment for partial course credit. All participants completed both blocked and interleaved conditions, using different categories. Forty-eight participants completed the simultaneous condition and 48 others the successive condition. Twenty-six participants (14 in the simultaneous group and 12 in the successive group) did not reach the criterion of 95% correct responses during the secondary task during the study phase in one or both schedule conditions (see below for the details) and were excluded from further analyses.

Stimuli and procedure

This experiment followed a procedure similar to that of Experiments 1 and 2, except for the following changes. Only the low-similarity category set from Experiments 1 and 2 was

used in this experiment. During the study phase, in both the simultaneous and successive conditions participants were not required to try to “guess” the correct category assignment. On each trial a new stimulus was presented in the center of screen (successive condition) or on the right side of the screen (simultaneous condition) for 2,500 ms. This timing is equivalent to the feedback and study times from the previous experiments added together. During this time participants saw the correct assignment of the category above the exemplar. Participants were asked to study the stimulus and correct category assignment in order to perform well in a subsequent categorization task.

Additionally, on each trial, after studying the stimulus three buttons with the name of each of the three categories would replace the stimulus and correct category label on the screen and participants had to click the name of the category they had just seen. This was a secondary task included to guarantee that participants were paying attention to the category study task. The task is nonetheless a passive category learning task because the participants simply need to repeat the category shown to them. Participants completed two study blocks, each composed of 72 trials.

During the test phase after a brief presentation of the stimulus (500 ms as in the previous experiments) participants had to click the button on the screen indicating the name of the guessed category. No feedback was given during the test phase. Participants completed a total of 48 test trials.

The labels used for the six categories (three presented interleaved and three presented blocked) were “beme,” “kipe,” “vune,” “coge,” “zade,” “tyfe” (Hendrickson, Kachergis, Fausey, & Goldstone, 2012). These are novel English-like words that do not share their initial letter and are equated for number of syllables and final sound. Category-label assignment was randomly determined at the beginning of the experiment for each participant.

Results and discussion

Data from the secondary task were analyzed, and participants that failed to repeat the correct category label on more than 5% (approximately seven trials in the total of 144) of the total number of trials were excluded from further analysis.

The results from the test phase are presented in Fig. 8. As can be seen in the graph, no main effect of presentation mode (simultaneous vs. successive) on overall accuracy is apparent. However, as we have seen before, performance was better for studied items than for novel ones. Of particular interest, blocked study improved performance during test, particularly for the successive group.

A mixed ANOVA with Presentation Mode (simultaneous vs. successive) as a between-subjects factor and Schedule of Presentation and Type of Stimuli (novel vs. studied) as within-subjects factors confirmed this interpretation. Performance

was overall higher for studied stimuli, $F(1, 68) = 51.47$, $MSE = .014$, $p < .0001$. Moreover, we found a main effect of schedule of presentation, with better performance following blocked study than interleaved study, $F(1, 68) = 5.48$, $MSE = .045$, $p = .02$. No effect emerged of presentation mode, $F(1, 68) = 2.37$, $MSE = .17$, $p = .13$, or interaction between any of the variables (all p s $> .05$).

Overall, these results replicate the findings from Experiments 1 and 2 for low-similarity categories. Blocked study resulted in better generalization performance than did interleaved study, and this was the case both for successive and simultaneous presentations, although less pronounced for the latter. When two stimuli were presented simultaneously, reducing memory constraints and allowing for greater contrast, performance following blocked study did not seem to be promoted as compared with successive comparison.

However, performance in the interleaved condition seemed to improve slightly in the simultaneous, as compared to the successive, presentation condition. A t test comparing performance between the successive and simultaneous groups for the interleaved condition only confirmed this interpretation, $t(68) = 2.02$, $p = .05$, $d = .48$. These results are in agreement with the proposal that simultaneous presentation increases the relative salience of differences between categories, improving performance following interleaved study.

General discussion

Inductively learning the characteristics of concepts and categories takes place frequently. It is important to be able to generalize the knowledge acquired when learning a set of examples. In this work, we studied the interaction between the ways that categories are sequenced during study and the structure of the categories being taught for inductive learning and generalization optimization.

The results from the experiments presented here show that three different factors interact during category learning to promote improved generalization: (1) the structure of the categories being learned (high vs. low similarity); (2) temporal presentation of exemplars (successive vs. simultaneous); and (3) the schedule of study of the categories (interleaved vs. blocked). Considering any one of these factors in isolation yields only a partial appreciation of the problem—all three factors work together to systematically guide learning, resulting in different generalization trends. This interaction may be parsimoniously explained by how learners must allocate their attention in order to meet the unique challenges of different learning tasks.

Study sequence and category structure

The experiments presented here show that interleaving categories during study promotes later generalization when high-

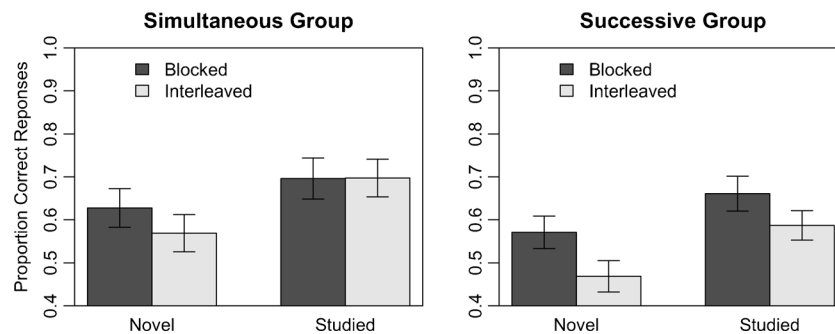


Fig. 8 Results for the generalization phase of Experiment 3. Error bars indicate standard errors of the means. Chance-level performance in this task was .33

similarity categories are being learned. For low-similarity categories, blocked study of categories resulted in better generalization in some situations.

In line with Goldstone (1996), we propose that interleaving categories allows participants to identify the features that distinguish among the categories, whereas blocked presentation promotes the identification of the features that are common among stimuli within the same category. This dichotomy is the result of the same principle: The opportunity to compare and contrast the properties of successive objects, which will emphasize different features in different situations.

Rapid alternation of categories allows participants to identify differences between categories, which will be particularly beneficial if those differences are hard to detect, as in the case of the stimuli in the high-similarity set used here, and the artists' styles or bird species used in previous studies (Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim et al., 2011). Infrequent alternation of categories, on the other hand, will facilitate participants' identification of the commonalities within each category, which is particularly beneficial if variability is high among the members of the categories, such the ones in the low-similarity set used in the present work.

Study sequence and simultaneous presentations

Varying the properties of the categories being learned changes which presentation order is more beneficial for later generalization. If the advantages of blocking and interleaving are due to opportunities to compare and contrast objects that are temporally close to each other, then presenting objects simultaneously in pairs may confer category learning benefits. Simultaneous presentation of two objects facilitates their comparison because it does not require one object (and its correct category assignment) to be stored in short-term memory.

The results from Experiment 2 show that simultaneous presentation indeed resulted in good generalization performance for the interleaved condition when learning high-similarity categories. However, simultaneous presentation in

the study of low-similarity categories resulted in equivalent generalization between the interleaved and blocked conditions (Exp. 2) although when in Experiment 3 simultaneous and successive presentations during the study of low-similarity categories were directly compared no difference in performance was found between the two. These are somewhat surprising results given the advantage for simultaneous presentation for high-similarity categories and previous research (Mundy et al., 2007, 2008; Vlach et al., 2012).

One possible reason for these results is that simultaneity increases the saliency of differences between objects belonging to different categories (Lipsitt, 1961; MacCaslin, 1954; Rieber, 1966). As we noted in the introduction, category learning can take the form of finding similarities among objects belonging to the same category or differences between objects coming from different categories. The latter process might be emphasized by simultaneous presentation and an expectation from participants that what one needs to do in these kind of categorization tasks is find differences between categories. However, for low-similarity categories, this strategy will not result in improved performance when associated with blocked study and might deter participants from implicitly finding the relevant within-category similarities. It did, however, improve performance for interleaved study, reducing the magnitude of the benefit of blocked study when associated with simultaneous presentation, which indicates that simultaneous presentation indeed increases the saliency of differences between the objects in the present work.

Sequencing effects in category learning: Sequential comparisons and attention allocation

The role of allocating one's attention during category learning has been highlighted before in different models (Kruschke, 1992; Love, Medin, & Gureckis, 2004; Minda & Smith, 2002; Nosofsky, 1986) and the use of eye tracking technology has made it possible to study the patterns of overt trial-by-trial, or even within-trial, attention. For example, Blair, Watson, Walshe,

and Maj (2009) have demonstrated that in a categorization task different stimuli can elicit different patterns of attention allocation to their features. Additionally, previous research has also demonstrated that during category learning participants take into account information from only the previous few trials to decide whether a stimulus belongs in one category or another (Jones, Love, & Maddox, 2006; Jones & Sieck, 2003; Stewart & Brown, 2004; Stewart, Brown, & Chater, 2002; Stewart & Chater, 2002).

As a theoretical framework for the sequencing effects presented here, we propose that during the presentation of objects, participants update their attentional target progressively toward the relevant features of the stimuli. To do this, learners take into account the information presented in the previous trial and the similarity relations between the previous object and the current object. If the previous trial consisted of an object in one category and the current trial consists of another object in a different category, participants' attention will be directed toward the *differences* between the two objects, by comparing the current object to the previous one (or their recollection, in the case of successive presentations). Conversely, if the two objects come from the same category, learners will attend to *similarities* between the objects.

This would result in attention being directed to the hard-to-find differences when high-similarity categories are interleaved and the substantial similarities when they are blocked. In much the same way, when interleaving low-similarity categories, attention will be directed toward the substantial differences, whereas if they are blocked, the hard-to-find similarities will receive more attention. This framework can parsimoniously account for the differential weighting of similarities and differences in the blocked and interleaved schedules, respectively, and the different advantage of each for different category learning situations.

Moreover, the framework we propose captures the results from the simultaneous presentation of high-similarity categories. Simultaneous presentation reduces the noise associated with using an imperfect recollection of the previous object, resulting in better comparison and contrast of the objects. However, it might also introduce new biases toward the differences between objects of different categories. Another possibility is that in order to identify similarities between highly different objects belonging to the same category, partial forgetting of the object properties is actually beneficial, helping participants to disengage their attention from the highly salient differences. This remains an open question for future research.

Conclusions

In this article, we showed, for the first time in a systematic way, the intricacy of the relations between “same” versus “different” category comparisons, interleaved versus blocked

study, and successive versus simultaneous presentations. Furthermore, we proposed trial-by-trial, category-specific attention allocation as the basis for the effects of the schedule of presentation on category learning.

In appreciating the benefits of one schedule of study over another, it is important to keep in mind that not all concept learning takes place by identifying discriminating features among categories. For example, sometimes it is possible to create an absolute characterization of a category in terms of its prevalent features, regardless of their discriminative values (Markman & Ross, 2003). Furthermore, in other situations, memorizing instances might be a highly useful strategy. The results presented here show that there might not be one universal answer to the question of whether an instructor should interleave or block information so that the learner acquires the knowledge and is able to generalize it more efficiently. Best sequencing practices will depend on the nature of the categories being ordered and other contextual factors.

Author Note Research supported in part by National Science Foundation REESE Grant No. 0910218 and Department of Education IES Grant No. R305A1100060. P.F.C. was also supported by a Fulbright Research Fellowship and by Graduate Research Fellowship SFRH/BD/68554/2010 from the Portuguese Foundation for Science and Technology (FCT), co-financed by the European Social Fund. The authors thank Caitlin Fausey and Drew Hendrickson for helpful discussion and suggestions, and Rachel Selonick, Kelly Rapp, Spenser Bengel, and Abigail Kost for their assistance with data collection.

References

- Andrews, J. K., Livingston, K. R., & Kurtz, K. J. (2010). Category learning in the context of co-presented items. *Cognitive Processing, 12*, 161–175.
- 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. doi:10.3758/s13421-012-0272-7
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1196–1206. doi:10.1037/a0016272
- Delaney, P. F., Verhoeven, P. P. J. L., & Spigel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 63–147). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(10)53003-2
- Dwyer, D. M., Mundy, M. E., & Honey, R. C. (2011). The role of stimulus comparison in human perceptual learning: Effects of distractor placement. *Journal of Experimental Psychology: Animal Behavior Processes, 37*, 300–307. doi:10.1037/a0023078

- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York, NY: Teachers College, Columbia University.
- Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 397–417.
- Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, 12, 20–30.
- Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, 14, 487–513. doi:10.1016/S0885-2014(99)00016-7
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, 24, 608–628. doi:10.3758/BF03201087
- Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, 107, 280–290. doi:10.1016/j.jecp.2010.04.017
- Hammer, R., Bar-Hillel, A., Hertz, T., Weinsall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research*, 1225, 102–118. doi:10.1016/j.brainres.2008.04.079
- Hendrickson, A. T., Kachergis, G., Fausey, C. M., & Goldstone, R. L. (2012). Re-learning labeled categories reveals structured representations. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1668–1673). Austin, TX: Cognitive Science Society.
- Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1388–1393). Austin, TX: Cognitive Science Society.
- Homa, D., & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 351–359. doi:10.1037/0278-7393.1.4.351
- Hull, C. L. (1920). Quantitative aspects of evolution of concepts. *Psychological Monographs: General and Applied*, 28, 1–86.
- Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 316–332. doi:10.1037/0278-7393.32.3.316
- Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 626–640. doi:10.1037/0278-7393.29.4.626
- Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, 26, 97–103. doi:10.1002/acp.1801
- Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23, 1297–1317. doi:10.1002/acp.1537
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19, 585–592. doi:10.1111/j.1467-9280.2008.02127.x
- Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, 25, 498–503. doi:10.1037/a0017807
- Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797–2822.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44. doi:10.1037/0033-295X.99.1.22
- Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology*, 51, 239–243. doi:10.1037/h0040295
- Lee, E. S., MacGregor, J. N., Bavelas, A., & Mirlin, L. (1988). The effects of error transformations on classification performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 66–74. doi:10.1037/0278-7393.14.1.66
- Lipsitt, L. P. (1961). Simultaneous and successive discrimination-learning in children. *Child Development*, 32, 337–347.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309–332. doi:10.1037/0033-295X.111.2.309
- MacCaslin, E. (1954). Successive and simultaneous discrimination as a function of stimulus-similarity. *American Journal of Psychology*, 67, 308–314.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129, 592–613. doi:10.1037/0033-2909.129.4.592
- Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, 16, 1050–1057. doi:10.3758/PBR.16.6.1050
- Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 275–292. doi:10.1037/0278-7393.28.2.275
- Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 237–242.
- Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology: Animal Behavior Processes*, 33, 124–138. doi:10.1037/0097-7403.33.2.124
- Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2008). Superior discrimination between similar stimuli after simultaneous exposure. *Quarterly Journal of Experimental Psychology*, 62, 18–25.
- Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology*, 131, 5–15.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–61. doi:10.1037/0096-3445.115.1.39
- Proctor, R. W. (1980). The influence of intervening tasks on the spacing effect for frequency judgments. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 254–266. doi:10.1037/0278-7393.6.3.254
- Rieber, M. (1966). Role of stimulus comparison in children's discrimination learning. *Journal of Experimental Psychology*, 72, 263–270.
- Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, 39, 406–412. doi:10.3102/0013189X10374770
- Sandhofer, C. M., & Doumas, L. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development*, 9, 194–221. doi:10.1080/15248370802022639
- Sims, C. E., & Colunga, E. (2010). When comparison helps: The role of language, prior knowledge and similarity in categorizing novel objects. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 339–404). Austin, TX: Cognitive Science Society.
- Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning—effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1251–1263.
- Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science*, 19, 1169–1177.
- Stewart, N., & Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. *Journal of Experimental*

- Psychology: Learning, Memory, and Cognition*, 30, 416–430. doi:10.1037/0278-7393.30.2.416
- Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 3–11. doi:10.1037/0278-7393.28.1.3
- Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 893–907.
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24, 837–848.
- Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 246–254. doi:10.1037/a0025260
- Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109, 163–167. doi:10.1016/j.cognition.2008.07.013
- 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
- Whitman, J. R., & Garner, W. R. (1963). Concept learning as a function of form of internal structure. *Journal of Verbal Learning and Verbal Behavior*, 2, 195–202.