Psychon Bull Rev
DOI 10.3758/s13423-014-0676-4

BRIEF REPORT

The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study

Paulo F. Carvalho · Robert L. Goldstone

Abstract Research on how information should be studied during inductive category learning has identified both interleaving of categories and blocking by category as beneficial for learning. Previous work suggests that this mixed evidence can be reconciled by taking into account within- and between-category similarity relations. In this article, we present a new moderating factor. Across two experiments, one group of participants studied categories actively (by studying the objects without correct category assignment and actively figuring out what the category was), either interleaved or blocked. Another group studied the same categories passively (objects and correct category assignment were simultaneously provided). Results from a subsequent generalization task show that whether interleaved or blocked study results in better learning depends on whether study is active or passive. One account of these results is that different presentation sequences and tasks promote different patterns of attention to stimulus components. Passive learning and blocking promote attending to commonalities within categories, while active learning and interleaving promote attending to differences between categories.

Keywords Interleaving · Blocking · Category learning · Comparison · Inductive learning

The method by which information is presented can substantially affect learning. Changing the way in which information is presented changes not only what is learned (Schyns, Goldstone, & Thibaut, 1998), but also how well it is learned (Goldstone, 1996). One example is the order in which instances are presented in a study session and the effect this has for inductive learning. Kornell and Bjork (2008) demonstrated that if participants are given study examples of paintings from several artists interleaved, participants' later memory and generalization is substantially improved, when compared with presenting each artist in a separate block (for similar results with different stimuli and tasks, see Wahlheim, Dunlosky, & Jacoby, 2011; Taylor & Rohrer, 2010).

Notwithstanding the clear benefit of interleaving in some situations, there have also been demonstrations of the advantage of blocking for category learning. For example, Goldstone (1996) compared interleaved and blocked study schedules. On each trial, participants had to classify a novel complex stimuli created by distortion of one of two prototypes. Categorization performance was better for the blocked condition. Goldstone argued that the relative difficulty in finding the common features shared by members of the same category contributed to this advantage (for further evidence of blocking advantages using different kinds of tasks and stimuli, see Carpenter & Mueller, 2013; Higgins & Ross, 2011; Kurtz & Hovland, 1956; Whitman & Garner, 1963). Moreover, when given the opportunity to choose, the overwhelming majority of learners decide to block their study (Braithwaite, Carvalho, de Leeuw, Motz, & Goldstone, 2014; Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013).

Given this mixed evidence about the best way to sequentially present information for optimal learning and people's preference for blocked study, an important question is: What conditions yield an advantage for interleaving, as compared with blocking?

This question has received some attention in recent years. For instance, Carvalho and Goldstone (2014) showed that when categories in which the exemplars share few similarities within and between categories (low similarity categories) were studied, blocked presentation resulted in improved subsequent generalization performance. This pattern was reversed

Electronic supplementary material The online version of this article (doi:10.3758/s13423-014-0676-4) 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

Published online: 02 July 2014

for high-similarity categories, categories in which all the exemplars share a high level of similarity with other exemplars in the same category, as well as exemplars in different categories (for similar results with category discriminability, see Zulkiply & Burt, 2013).

Carvalho and Goldstone (2014; see also Goldstone, 1996) have proposed that interleaving categories allows participants to identify the features that distinguish between elements of each category, while blocked presentation promotes the identification of features that are characteristic among stimuli within a single category. This dichotomy is the result of the same principle: the selective emphasis of categorization-relevant features during comparison of sequentially presented objects.

In the case of interleaved presentation, differences between objects belonging to different categories will be emphasized, while for blocked presentation, similarities among objects belonging to the same category will be emphasized. This same process will result in improved or hindered learning depending on whether it is more important in a task for similarities or differences to be learned. One possible way in which category learning could shift from an emphasis on differences toward an emphasis on similarities is by changing the similarity relations within and between categories (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013).

Active and passive category learning

Going beyond study schedule of presentation, there are different ways in which category learning can take place. One way is by an active process of classifying an object into one of the possible categories and then receiving feedback on the categorization. Another way is to passively study the exemplars along with their correct category assignment. Thus, active study can be seen as a type of classification learning in which an object's features are used to predict a categorization, while passive study can be seen as a type of inference learning, in which given a label and a set of characteristics, the relevant feature (the one more correlated with the label) needs to be inferred.

These two category acquisition tasks differ in a number of relevant properties. Importantly, they emphasize different properties of the category structure. It has been argued before (see, e.g., Markman & Ross, 2003; Yamauchi & Markman, 2000) that classification learning involves a process of discrimination and identification of the relevant discriminating feature(s), while inference learning involves acquiring information about how the features within a category are related and organized (this distinction might be related to the distinction between rule-based and information integration categorization in COVIS; see Ashby, Alfonso-Reese, Turken, & Waldron, 1998). Some evidence for this distinction is the finding that classification learners are more sensitive to diagnostic features, while inference learners are sensitive to the prototypical features (Chin-Parker & Ross, 2004). Additionally, it has been demonstrated before that unsupervised learning situations create a strong bias for rule-based categorizations (Ashby, Queller, & Berretty, 1999). It could be argued that the passive category learning described previously differs from unsupervised learning because participants get information about the correct category assignment. However, both require identifying regularities among features, similar to non-error-driven learning (e.g., Soto & Wasserman, 2010).

This critical difference between the two types of tasks can be paired with interleaving or blocking to result in better or worse learning. One proposal is that in an active learning task participants would try to discriminate between categories and, according to Carvalho and Goldstone's (2014) proposal, interleaving provides them with a sequence of exemplars that helps finding the discriminative features. Conversely, participants in a passive learning task would try to create a positive characterization of each category, and blocking provides them with the sequence of exemplars that helps them find the common, unique features, of each category. This is not to say that learning would not be possible with, for example, a blocked sequence in active study, but this sequence would not optimize discrimination learning and, thus, would be less efficient for learning.

In this article, we aim to extend previous evidence for a comparison and attentional mechanism as the unifying processes behind both blocked and interleaved study advantages. To that end, we manipulate the properties of the study session that affect attention allocation. In essence, we propose that interleaving and blocking result in different attentional patterns of encoding and whether one or the other is more beneficial for learning is the result of the match between the attentional encoding and task demands.

Across two experiment using different types of categories, one group of participants completed a passive study session associated with both interleaved and blocked presentations. The other group of participants completed an active study session, while keeping all other aspects of the task constant between the two groups.

Experiment 1

Participants studied a set of six categories, three presented interleaved and the remaining three presented blocked. Critically, for one group of participants, the study session was set up as a passive learning task. Participants studied each object for a short period of time during which the correct category assignment was also presented on the screen. For the other group of participants, the categories were studied in an active learning task in which the correct category assignment was not immediately available. Both groups completed the same generalization task afterward.

Method

Participants

Eighty-two undergraduate students at Indiana University volunteered to participate in return for partial course credit. Forty-four participants were randomly assigned to the passive learning group, while the remaining 38 were assigned to the active learning group. Data from 7 participants in the passive learning group were excluded from analyses due to failure to repeat the label of the object just presented on more than half of the total number of study trials (see below for details). All participants in the active learning group reached the criterion of 34% correct responses during study, and their data were kept for analyses. All participants completed both an interleaved and a blocked study condition.

Apparatus and stimuli

In this experiment, stimuli were "Fribble" objects (Williams, 1997). Three of the categories were composed of very similar objects differing diagnostically only in one of their parts (see the top panel in Fig. 1). The other three categories were also very similar and differed diagnostically from each other only in one of their parts; however, they were substantially different from the other three categories (see the bottom panel in Fig. 1). Random variation existed in each of the categories but was the same across the three categories in each set (see Fig. 1). Each category set was randomly selected for each participant to be used for the blocked or interleaved study condition. From each category, four items were randomly selected to be included during the study phase, while the remaining were used as novel items during the test phase only (see the Design and Procedure section for details).

Each category was given a unique label that perfectly predicted the presence of the unique feature that defined that category. At the start of the experiment, one label was randomly picked for each category from the following pool: *beme, kipe, vune, coge, zade, and tyfe*.

Design and procedure

This experiment had two conditions manipulated within subjects (schedule of study: blocked category study vs. interleaved category study) and two conditions manipulated between subjects (study type: active vs. passive). Each of these four conditions was composed of a study and test phase.

Study phase For the passive study group, during this study phase, participants were presented with a stimulus in the center of the screen, along with the correct category assignment above the object, for 2.5 s. Immediately after the presentation of the stimulus, three buttons were shown on the screen corresponding to the three possible category names for that study session. The participants' task was to press the button corresponding to the category of the object they just saw. This task was introduced to ensure participants' attention to the task and to equate the active and passive learning situations for the presence of a motor response. However, note that in this group, participants simply needed to repeat the category they had just seen. There was no need for participants to learn a categorization rule. The mapping between the position of the buttons on the screen and the label was randomly shifted each trial.

For the active learning group, participants were presented with a stimulus for 500 ms without its label. After the stimulus was removed, three buttons were shown on the screen, and the participants had to choose the category assignment for that stimulus. After the participants' response, the stimulus as well as response feedback and the correct category assignment were shown on the screen for 2 s. The mapping between the position of the buttons on the screen and the label was also randomly shifted each trial.

For both groups, a 1-s intertrial interval followed the trial, and then the next trial began. In the blocked condition, the categories presented alternated 25% of the time, while in the interleaved condition, they alternated 75% of the time, similar to the procedure used in previous work (Carvalho & Goldstone, 2014; Goldstone, 1996). For example, one representative sequence of the categories A, B and C might be "AAABCACBCBCBCBABAACBBCCC" for the interleaved condition, while a representative sequence for the blocked condition would be "AAABBBBBBBBBBCCCC CBBBCAA." The particular order of trials was pseudorandomly assigned by the computer at the beginning of each study block, following the criteria just described. Thus, there was no preset structure of alternation/repetition in the experiment, and every participant received a different sequence of items.

Each study phase was composed of four blocks for both groups of participants, and the entire study phase took approximately the same amount of time for each group. Each block had 24 trials (two repetitions of four exemplars of each of the three categories). After the fourth block of study, a new set of instructions was presented on the screen, and the second phase began. Each participant completed two sets of study and test phases (one for each schedule of presentation).

The two schedule-of-study conditions (blocked vs. interleaved) differed only in the frequency of category change during study and the category labels. Which condition was presented first was counterbalanced across participants, and the allocation of the stimuli to each category and condition was randomized across participants.

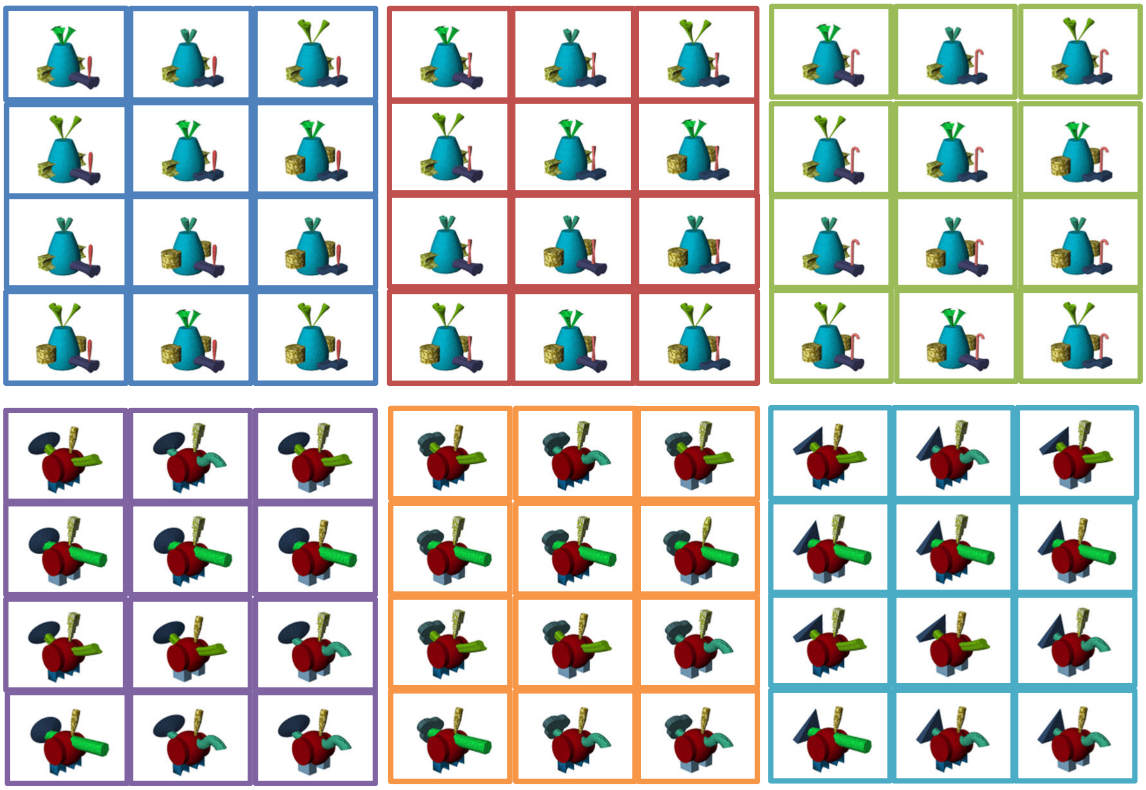

Fig. 1 Exemplars of each of the six categories used. The top row constitutes one set of categories, and the bottom row another set. The color of the squares around the images indicates the category. Participants were presented with only one object at a time, in a white background with no square borders around it

Test phase This second phase was a transfer task during which 36 stimuli were shown in random order: the 12 objects participants studied during the study phase and 24 new stimuli. The new stimuli were similar to the studied stimuli, with new instantiations of the unique features (i.e., the unique feature presented with different nondiagnostic 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, by pressing a button on the screen. After a 1-s intertrial interval, a new trial would begin. The mapping of category names to button locations was randomly assigned each trial. No feedback was provided during this phase. Each test phase followed the respective study phase.

Results and discussion

Because the dependent variable of interest in this work (accuracy) is proportional in nature, accuracy data were first submitted to an empirical *logit* transformation for this and the subsequent experiment.

Performance in the transfer task for participants in the active and passive study conditions and for each study presentation condition is presented in Fig. 2. We started by analyzing the possibility that one of the conditions resulted in overall better learning, using a mixed ANOVA model including order of conditions as a between-subjects factor. Which condition (interleaved or blocked) participants completed first did not have a significant effect on transfer performance, $F(1, 71) < 1$, while, overall, participants' performance was better in the second condition than in the first, $F(1, 71) = 6.04$, $p = .02$, $\eta_G^2 = .04$.

We conducted a mixed ANOVA with presentation schedule (blocked vs. interleaved) and type of item (studied vs. novel) as within-subjects factors and study group (passive vs. active) as a between-subjects factor, not taking into account the order of conditions. This analysis revealed no main effects of presentation schedule, $F(1, 73) < 1$, study group, $F(1, 73) < 1$, or type of item, $F(1, 73) < 1$. However, the critical crossover interaction between schedule of presentation and study group was reliable, $F(1, 73) = 4.22$, $p = .04$, $\eta_G^2 = .03$. More specifically, the relative advantage of each study schedule varies depending on the type of study session.

In sum, the results show that whether interleaving examples of several concepts or blocking examples by category is beneficial is a function of the training task's implicit demands. Importantly, this interaction is seen for both novel and studied items, indicating that it is not limited to better memory of the individual exemplars presented.

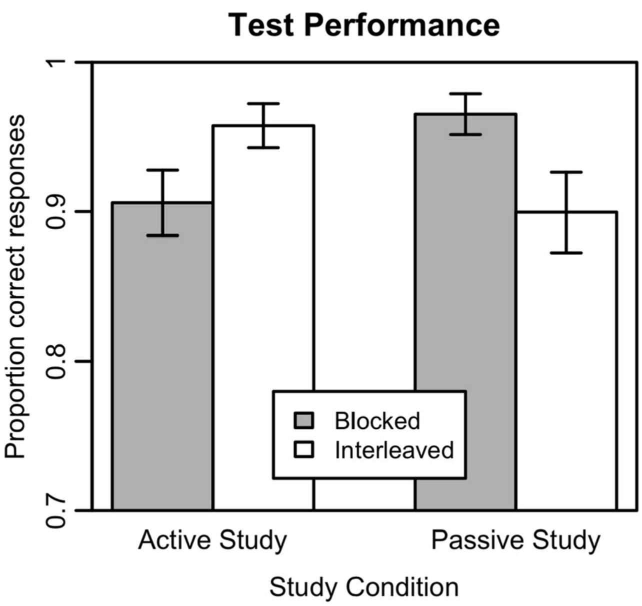

Fig. 2 Results for the test phase of Experiment 1. The nontransformed data were used to create this plot. Error bars indicate standard errors of the means. Chance level performance in this task is .33

Experiment 2

This second experiment provides a conceptual replication as well as an expansion of the results of Experiment 1. One open question resulting from Experiment 1 is whether the results presented are unique to the category type used or would extend to categories with different structures (a manipulation previously shown to affect the advantage of different study schedules; see Carvalho & Goldstone, 2014). In this experiment we replicate the procedure of Experiment 1 using categories with high between-category similarity but low within-category similarity. This similarity structure retains the critical difficulty in finding differences between categories while also including greater difficulty in finding similarities within categories, allowing for a greater range of categorization strategies. Moreover, the categories used here required noticing several relevant properties for correct category assignment, unlike in Experiment 1. If the interaction seen in Experiment 1 is the result of task-appropriate processing between the study schedule and the type of task, the interaction between study group and schedule of presentation should remain.

Method

Participants

One hundred twenty-nine undergraduate students at Indiana University volunteered to participate in this study in return for partial course credit. Sixty-five participants were randomly assigned to the active group, while the remaining 64 were assigned to the passive group. Using the same exclusion criteria as in Experiment 1, data from 7 participants in the passive group and from 2 participants in the active group were excluded.

Apparatus, stimuli, and procedure

The stimuli used for this experiment were a subset of *Ziggerins* created by Wong, Palmeri, and Gauthier (2009). Each category was composed of 12 stimuli and was defined by a property shared by all its members (a combination of cross-sectional shape, concavity, and shape of extremities; see Fig. 3). The same apparatus and procedure as in Experiment 1 were used in this experiment.

Results and discussion

We conducted the same set of analyses for this experiment as for Experiment 1. Similar to Experiment 1, which condition was completed first did not have an impact on performance during test for either group, $F(1, 116) < 0$. In general, participants performed better on whichever condition was second, relative to first, during test, $F(1, 116) = 1.88$, $p < .0001$, $\eta_G^{\mathbf{2}} = .04$.

The results from the transfer phase are reported in Fig. 4. Similar to Experiment 1, no reliable differences between novel and old items or interactions with any of the other variables were found (*ps* < .05). The results show an interaction between study group (passive vs. active) and presentation schedule (interleaved vs. blocked), $F(1, 118) = 4.59$, $p = .03$, $\eta_G^2 = .01$. Moreover, performance was, overall, better for the active group than for the passive group, $F(1, 118) = 10.27$, $p = .002$, $\eta_G^2 = .06$. Overall, these results replicate those of Experiment 1 and extend them to categories with a different structure.

General discussion

Determining how to order information so that learners can achieve the best learning outcomes is crucial for effective training. Here, we present further evidence that the way information is ordered impacts learning and that this influence is modulated by the demands of the study task—in the present case, whether learning is active or passive. We propose that the relative benefits of each study schedule are modulated by how well the attentional biases they promote during learning match the demands of the learning task. While both active and interleaved study encourage discrimination of the concepts being studied, passive and blocked study encourage creating an independent, stand-alone representation of the concepts by identifying the similarities among instances within each category. By this account, learners in the active study condition tend to look for features that discriminate between categories, and

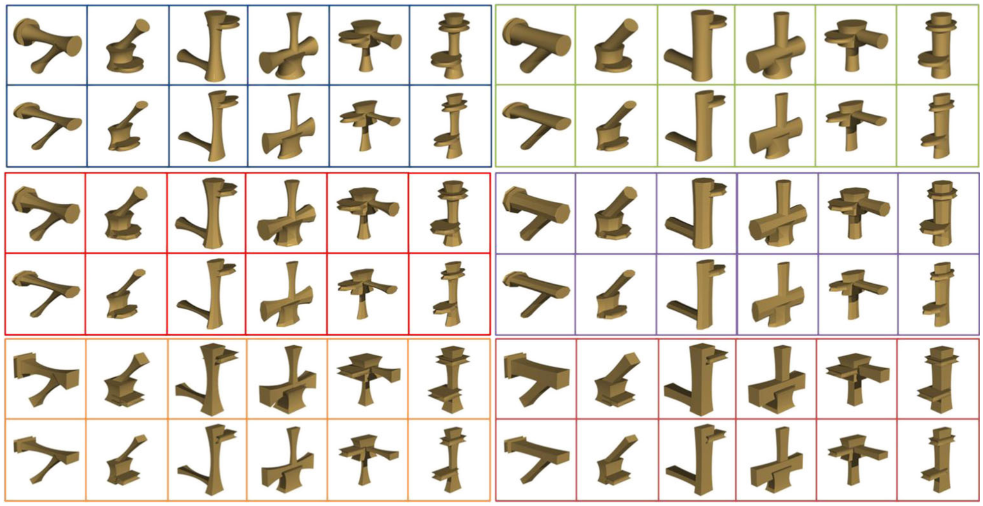

Fig. 3 Exemplars from each of the categories used in Experiment 2. The color of the squares around the images indicates the category. Participants were presented with only one object at a time, in a white background with no square borders around it

these features are easier to detect when categories frequently alternate. Learners in the passive study condition are more likely to look for features that consistently appear within one category's examples, and these features are easier to detect when categories rarely alternate. This proposal matches the direction of the interaction seen here.

In addition, the interaction between the type of study situation and the schedule of presentation of the exemplars had an effect for both studied stimuli and novel ones. This is an interesting result, suggesting that study benefits go beyond memorization of the whole exemplars. Very likely, participants succeeded at the categorization task by identifying the

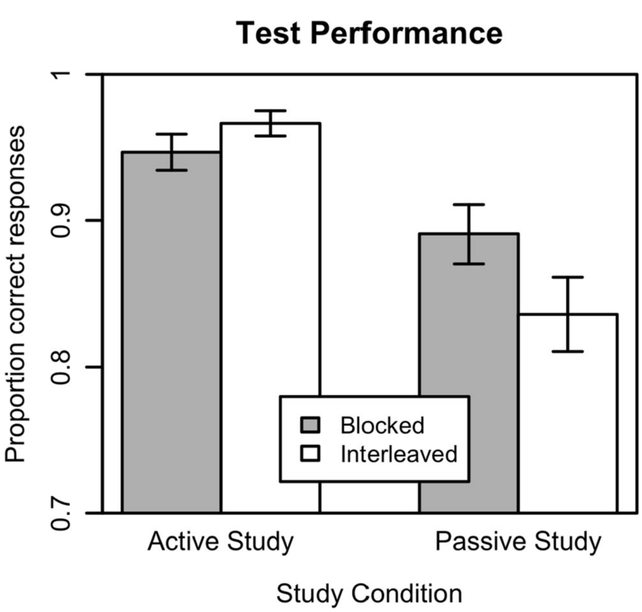

Fig. 4 Results for the test phase of Experiment 2. The nontransformed data were used to create this plot. Error bars indicate standard errors of the means. Chance level performance in this task is .33

defining parts for each category. These defining parts were instantiated identically for studied and novel objects, explaining why novel objects were not more difficult to categorize.

Moreover, with this type of category, the same set of features are emphasized by active and passive study (although by different processes, as discussed in the introduction), because they completely distinguish between categories and are highly associated with the category label. This might explain why we found similar results across two different types of categories. Future research will be needed to test whether the interaction between learning activity and presentation schedule generalizes to category structures not defined by a rule. It is possible that the particular direction of the interaction found here is related to the type of category used, in which a single feature identified the categories (contrary to previous passive study of categories in which interleaving was found to be more beneficial; e.g., Kornell & Bjork, 2008; Wahlheim et al., 2011). For example, it might be the case that information integration categories benefit from interleaved study and rule-based category learning benefits from blocked study. Given that unsupervised situations (similar to the passive study situation) promote rule-based categorizations (Ashby et al., 1999), the direction of the interaction might be the result of the use of rule-based categories sharing high between-category similarity but would be reversed with the use of information-integration categories. Thus, it is not our intention to imply that active study should always be interleaved and passive study should always be blocked but, rather, that the relative benefits of each study schedule will depend on the nature of the study task.

Overall, these results are consistent with the framework proposed by Carvalho and Goldstone (2014; see also

Goldstone, 1996) hypothesizing that participants compare successive objects and update attention to stimulus features as a result of these comparisons. In a series of experiments using an active study procedure, the authors demonstrated that blocked study resulted in better performance for categories with low within-category similarity, while interleaved study resulted in better performance for categories with high between-category similarity. These results are consistent with the present results for the active group because, in both experiments, the categories used shared a high level of between-category similarity.

Carvalho and Goldstone (2014) proposed that when studying a new exemplar in an inductive learning task, participants compare the properties of that object with the properties they recall from the previous ones. However, learners do not remember all the features from all the objects presented. Instead, when studying a new exemplar, learners weight more heavily the information presented in the immediately preceding instance. 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 preferentially directed toward the differences between the two objects. Conversely, if the two objects come from the same category, learners will tend to attend to similarities between the objects.

This framework can aptly account for the results presented here: Passive learning is facilitated by attending to similarities among objects belonging to the same category, while active learning is facilitated by attending to differences between objects in different categories. When the presentation order emphasizes these same factors, learning will be facilitated.

In conclusion, one of the most important contributions of the present work is the proposal that when deciding how to structure learning, one needs to take into account the entire learning situation and possible interactions between situational factors. So far, this has been demonstrated in the interaction between study schedule and category structure (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013) and, in the present work, relative to the type of study task. An important goal for future research would be to identify possible interactions between different sequences of exemplars during study and additional situational variables, such as individual differences (low vs. high knowledge of the category), developmental differences (young vs. older children), and at different time scales (weeks of study and retention interval, for instance).

Acknowledgments This research was supported in part by National Science Foundation REESE grant 0910218 and Department of Education IES grant R305A1100060. P.F.C. was also supported by Graduate Training Fellowship SFRH/BD/78083/2011 from the Portuguese Foundation for Science and Technology (FCT), co-sponsored by the European Social Found. The authors would like to thank the Percepts and Concepts Lab members for discussion and Spenser Benge, Abigail Kost, Alifya Saify, and Shivani Vasudeva for their assistance with data collection. The authors are also thankful to Bob McMurray and two anonymous reviewers for their feedback on earlier versions of the manuscript. "Freeble" stimuli images are courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, http://www.tarrlab.org/. "Ziggerin" stimuli images are courtesy of Alan Wong, The Chinese University of Hong Kong, http://ww2.psy.cuhk.edu.hk/~mael/Stimuli.html.

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*(3), 442–481.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*(6), 1178–99.

Braithwaite, D. W., Carvalho, P. F., de Leeuw, J., Motz, B. A., & Goldstone, R.L. (2014) Effectiveness of learner-regulated study sequence. Manuscript in preparation.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41*(5), 671–682.

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*(3), 481–495.

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 216–226.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*(5), 608–28.

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for? In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1388–1393). Austin: Cogntive Science Society.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585–92.

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology, 51*(4), 239.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*(4), 592–613.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837–848.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review, 20*(2), 356–363.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *The Behavioral and Brain Sciences, 21*(1), 1–17. Discussion 17–54.

Soto, F. A. & Wasserman, E.A. (2010) Error-driven learning in visual categorization and object recognition: A common-elements model. *Psychological Review, 117*(2), 349–381.

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*(5), 750–763.

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*(2), 195–202.

Williams, P. (1997). *Prototypes, Exemplars, and Object Recognition*. New Haven: Yale.

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science, 20*(9), 1108–1117.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 776–795.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41*(1), 16–27.