

Organized simultaneous displays facilitate learning of complex natural science categories

**Brian J. Meagher, Paulo F. Carvalho,
Robert L. Goldstone & Robert
M. Nosofsky**

Psychonomic Bulletin & Review

ISSN 1069-9384

Psychon Bull Rev

DOI 10.3758/s13423-017-1251-6



Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



BRIEF REPORT

Organized simultaneous displays facilitate learning of complex natural science categories

Brian J. Meagher¹ · Paulo F. Carvalho² · Robert L. Goldstone¹ · Robert M. Nosofsky¹

© Psychonomic Society, Inc. 2017

Abstract Subjects learned to classify images of rocks into the categories igneous, metamorphic, and sedimentary. In accord with the real-world structure of these categories, the to-be-classified rocks in the experiments had a dispersed similarity structure. Our central hypothesis was that learning of these complex categories would be improved through observational study of organized, simultaneous displays of the multiple rock tokens. In support of this hypothesis, a technique that included the presentation of the simultaneous displays during phases of the learning process yielded improved acquisition (Experiment 1) and generalization (Experiment 2) compared to methods that relied solely on sequential forms of study and testing. The technique appears to provide a good starting point for application of cognitive-psychology principles of effective category learning to the science classroom.

Keywords Category learning · Cognitive training

Introductory courses in geology often include a section on rock identification and classification. At a broad level, geologic scientists divide rocks into the categories *igneous*, *metamorphic*, and *sedimentary* (e.g., Tarbuck & Lutgens,

The data from the studies presented in this article are publicly available from: <https://osf.io/vh7pn>.

Electronic supplementary material The online version of this article (doi:10.3758/s13423-017-1251-6) contains supplementary material, which is available to authorized users.

✉ Robert M. Nosofsky
nosofsky@indiana.edu

¹ Department of Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405, USA

² Carnegie Mellon University, Adelaide, SA, Australia

2015). These scientific divisions are based on the manner in which the rocks were originally formed. However, these formation processes occurred eons ago and are not directly accessible. Thus, observers must integrate information from multiple perceptual dimensions of the rocks as a basis for classifying them. Indeed, college-level geology texts devote a great deal of space to describing and illustrating these perceptual characteristics of the different rock categories (e.g., Marshak, 2013; Tarbuck & Lutgens, 2015).

The task of learning to classify rocks is not an easy one. For example, Fig. 1 provides an illustration of a subset of rocks belonging to the igneous, metamorphic and sedimentary categories. As illustrated in the figure, rocks belonging to the same broad category (e.g., igneous) can be very dissimilar while simultaneously being highly similar to rocks from contrasting categories (e.g., metamorphic). Nosofsky, Sanders, Gerdman, Douglas, and McDaniel (2017) conducted a similarity-scaling study involving ten common subtypes of igneous, metamorphic and sedimentary rocks. The similarity structure of the rock categories was found to be highly disorganized and dispersed, much as is depicted in Fig. 1. Categories with these types of complex similarity structures have long been identified as difficult to learn.

In the work presented here we investigated how one might improve the acquisition of these complex natural categories. Guided by previous evidence from the study of category learning in cognitive psychology, our key idea was that a technique involving the organized, simultaneous display of the set of to-be-learned items might be highly beneficial.

The first source of evidence that motivated our hypothesis is that researchers have demonstrated that learning can be improved by studying two items of the same category simultaneously, as opposed to one at a time (Catrambone & Holyoak, 1989; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Hammer, Diesendruck, Weinshall, &

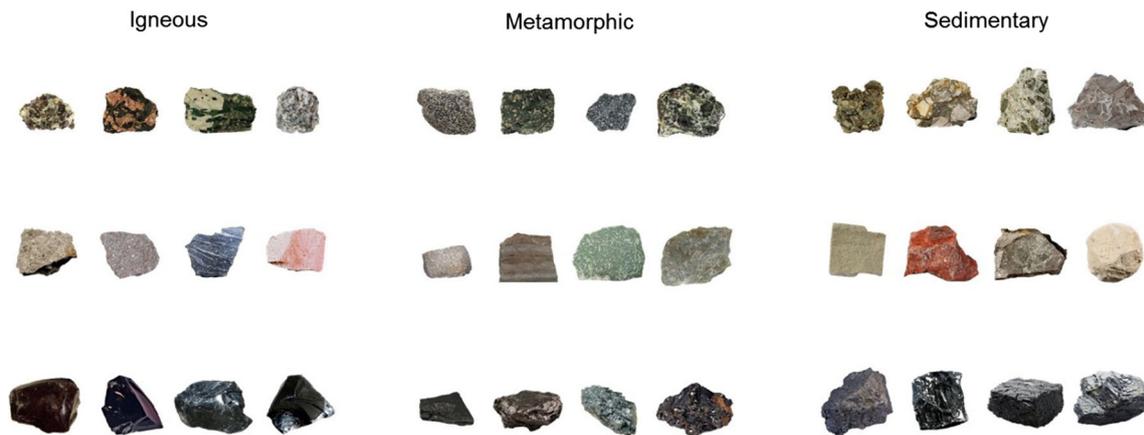


Fig. 1 Example of a Simultaneous display trial. Each of the three major columns corresponds to a different rock category (igneous, metamorphic, sedimentary). Each of the three rows corresponds to a separate similarity

cluster. Each grouping of four rocks within each row is a specific rock subtype (see Table 1)

Hochstein, 2009). For example, Gentner et al. (2003) found that teams of MBA students who simultaneously compared two training cases to identify a key “value-added” principle of effective negotiation were more likely to transfer the strategy to a novel negotiation than were teams who analyzed the same two cases separately. One advantage of the simultaneous comparison is that when the objects belong to the same category or schema, the properties that are common to members of the category are emphasized (Goldwater & Schalk, 2016; Kurtz, Miao, & Gentner, 2001; Markman & Gentner, 1993).

Reciprocally, simultaneous comparison of similar objects that belong to different categories is effective in highlighting the critical property that supports the distinction. Thus, presenting objects that belong to different categories simultaneously, or close in time, can also boost learning (Carvalho & Goldstone, 2014; Gentner & Markman, 1994). In the context of geology instruction, for example, Jee, Uttal, and Gentner (2008) demonstrated that studying two highly similar examples that differ only in terms of the presence of a “fault” improves learning of that concept. In a triad-presentation method, Andrews, Livingston, and Kurtz (2011) found that presenting members of three separate categories simultaneously on each trial led to better learning than presenting three members of the same category.

Learning of dispersed categories such as igneous, metamorphic and sedimentary rocks requires both the grouping of disparate items within categories as well as the detection of subtle features that discriminate between categories. A natural hypothesis, therefore, is that presenting multiple examples of the categories in a simultaneous organized display might be highly beneficial. In particular, an organized, simultaneous presentation of multiple examples of the categories combines both types of comparisons just described: it has the potential advantages of simultaneous study of items of the *same* category as well as simultaneous study of items of *different* categories. Furthermore, when all the information is available simultaneously, learners can choose how to organize

their study, which can improve learning (e.g., Gureckis & Markant, 2012). Simultaneous presentation also reduces the working memory load associated with sequential study: Because all the information is available at the same time there is no need to maintain in working memory the recently studied items in order to establish comparisons.

Although our key theme involves the use of an organized simultaneous display, we do not suggest that this technique will suffice if used on its own. In particular, compared to a sequential learning with feedback procedure, if learners are restricted to only observing the display, they will have no chance to experience the well-known benefits of testing during the learning process. For example, as researchers such as Roediger and Karpicke (2006) have made clear, the retrieval practice provided by testing yields benefits to long-term retention. In addition, testing allows learners to self-monitor their performance and keeps them engaged in the task. Accordingly, our proposal is not to use observational simultaneous displays on their own but rather to combine their use with that of sequential study-test procedures. Thus, as described in more detail in our [Method](#) section, to get started on our investigation, we chose to explore a particular training technique in which observers viewed an organized simultaneous display on some blocks and engaged in sequential study-test trials on other blocks.

We should emphasize that although we have focused on hypothesized benefits of organized simultaneous displays, there are also potential drawbacks to the technique. First, presenting multiple items simultaneously may increase cognitive load if observers attempt to deal with the entire display all at once. According to the cognitive load theory (Sweller, 2004), such increases have a detrimental effect on learning and transfer (e.g., van Merriënboer, Kirschner, & Kester, 2003). Another potential drawback involves the purely observational nature of the display. For example, in cases involving the learning of perceptual categories in which information from multiple dimensions needs to be integrated, observational

training has often been found to be inferior to techniques in which observers generate answers and then receive feedback (e.g., Ashby, Maddox, & Bohil, 2002; but see Levering & Kurtz, 2015, for an example of advantages of observational training). It is an open question whether the positive effects that we hypothesize are associated with the simultaneous displays will outweigh the potential drawbacks.

Experiment 1

Method

Subjects learned to classify varieties of rocks into the categories igneous, metamorphic and sedimentary. To mimic the real-world complex structure of these categories (see Nosofsky et al., 2017), we deliberately selected rock subtypes that were highly dispersed. There were three similarity clusters, each composed of three rock subtypes. Rocks within each cluster tended to be fairly similar to one another, but dissimilar to rocks from separate clusters. The dispersed category structure is created by selecting rock subtypes such that one subtype from each of the categories igneous, metamorphic, and sedimentary falls in each similarity cluster (see Fig. 1).

In the “Sequential” condition, all learning blocks involved a random sequence of rocks presented one at a time. On each trial, subjects attempted to classify an individual rock into its category and then received corrective feedback. In the “Simultaneous” condition, some of the sequential blocks were replaced with an organized, simultaneous display of all training items in a single screen. We organized the display such that all rocks within the same similarity cluster appeared in the same row of the display, and all rocks within the same category appeared in the same column (see Fig. 1). We used this organization with the goal of facilitating learners’ ability to establish comparisons both within and between categories.

Participants Sixty volunteers from the Indiana University community, each paid US\$12, participated in the study. Participants were randomly assigned to the sequential ($N = 29$) or the simultaneous ($N = 31$) conditions.

Materials The stimuli were photographs of rocks obtained from web searches. There were three subtypes of each of the high-level categories igneous, metamorphic, and sedimentary (see Table 1). (All subtypes were proper subsets of their respective high-level categories.) There were 12 tokens of each subtype. Each rock picture subtended a visual angle of approximately $7^\circ \times 7^\circ$ and was displayed on a white background. The experiment was conducted on PCs running MATLAB and the Psychophysics Toolbox (Brainard, 1997). For each participant, four of the 12 tokens of each subtype were randomly selected to serve as training stimuli, and two

Table 1 Rock subtypes of each category and similarity cluster

	Igneous	Metamorphic	Sedimentary
Cluster 1	Pegmatite	Amphibolite	Breccia
Cluster 2	Rhyolite	Quartzite	Dolomite
Cluster 3	Obsidian	Anthracite	Bituminous Coal

were randomly selected to serve as novel transfer stimuli. Thus, for each participant, there were totals of 36 training items and 18 novel transfer items.

Procedure Participants completed a learning phase followed by a transfer phase. During the learning phase participants completed six blocks, each containing 36 items (four training tokens of each of the nine rock subtypes).

The structure of the two conditions is shown in Table 2. In the Sequential condition, all six blocks were sequential-training blocks. Within each sequential block, the 36 training items were presented in a random order. Each trial lasted 7 s. On each trial, participants saw a rock token in the center of the screen and classified it into one of the three rock categories – igneous, metamorphic, or sedimentary – by pressing the “I,” “M,” or “S” keys, respectively (see Fig. 2). Participants had 5 s to respond; if they took longer, in the final 2 s they saw the phrase “You took too long!” followed by the correct category label. If the participant responded before the deadline, feedback was displayed for the remainder of the trial, and consisted of the phrase “Correct!” or “Incorrect” along with the correct category label. The rock was displayed throughout the entire 7-s period. There was a 1-s inter-trial interval.

The Simultaneous condition was the same as the Sequential condition except that during Blocks 1, 3 and 5, subjects were instead presented with an organized simultaneous display of the rocks. All 36 training tokens were presented simultaneously in a 3×12 grid (see Fig. 1). The display was organized such that all igneous rocks appeared on the left of the screen, all metamorphic rocks in the middle, and all sedimentary rocks on the right. Rocks in the first similarity cluster were all displayed in the top row, rocks in the second similarity cluster in the middle row, and rocks in the third similarity cluster in the bottom row. The first, second, and third group of four tokens in each row each corresponded to a particular subtype (identified in Table 1). Extra space was placed between subtypes of the different categories for emphasis. The order in which the tokens of each subtype appeared on the screen was randomized for each participant, but remained the same throughout the experiment. The headings “Igneous,” “Metamorphic,” and “Sedimentary” were displayed over the appropriate columns of rocks. Participants were informed that they would view the rocks for a total of 4 min and 12 s. Two-min, 1-min, and 30-s warnings were given to participants at the bottom of the screen. Note that the total time of the

Table 2 Schematic representation of the procedure of Experiment 1

Condition	Training						Transfer
	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	
Simultaneous	1 Sim Display Trial	36 Seq Trials	1 Sim Display Trial	36 Seq Trials	1 Sim Display Trial	36 Seq Trials	108 Transfer Trials
Sequential	36 Seq Trials	36 Seq Trials	36 Seq Trials	36 Seq Trials	36 Seq Trials	36 Seq Trials	

Note: Sim (Simultaneous) Display Trials are trials in which 36 tokens (12 from each category of rock) are presented simultaneously on the screen. Seq (Sequential) Trials are trials during which one rock is presented at a time on the screen, participants are asked to classify it and are then given feedback on their response

simultaneous display was identical to the total display time of the 36 rocks in each sequential block. For both conditions, at the end of each Sequential block, participants were informed of their overall percentage correct in that block.

During the transfer phase, the full set of 54 rock tokens (36 old tokens + 18 new tokens) was presented twice in a fully randomized order. On each trial, participants were shown a rock in the center of the screen and asked to classify it by pressing one of the keys “I,” “M,” or “S.” There was no time limit for this response. No feedback was provided; the experiment instructions explained to the subject that the word “Okay” would instead be displayed for 1 s following each response. There was a 1-s inter-trial interval. At the end of the transfer block, participants were informed of their percentage of correct responses. All participants were tested privately in individual sound-attenuated cubicles.

Results

In this and the subsequent experiment we started by analyzing the raw data for potential outliers (participants whose scores differed substantially from those of the sample). We defined outliers by investigating the boxplots of the sample for each condition and experiment. Participants whose scores were 1.5 standard deviations above the upper quartile or 1.5 standard deviations below the lower quartile were excluded from analysis (see [Supplementary Materials](#)). Two participants were excluded from the training-phase analysis (one in each



Igneous, Metamorphic, or Sedimentary?

Fig. 2 Example of a sequential display trial

condition) and five from the transfer-phase analysis (three in the simultaneous condition and two in the sequential condition).¹

Training phase We analyzed the results from sequential blocks 2, 4, and 6 (i.e., the blocks in which subjects from both conditions provided responses). As can be seen in the left panel of Fig. 3, participants performed better in the Simultaneous condition than in the Sequential condition in all three blocks. We confirmed this observation by conducting a 2×3 mixed-model ANOVA with condition (Simultaneous vs. Sequential) as a between-subjects variable and block (2 vs. 4 vs. 6) as a within-subjects variable. As expected, participants improved during the learning phase, $F(2, 112) = 239.89$, $p < .0001$, $\eta_G^2 = 0.479$. More importantly, participants were more accurate in the simultaneous condition than in the sequential condition, $F(1, 56) = 8.32$, $p = .006$, $\eta_G^2 = 0.104$. The interaction between condition and block was not significant, $F(2, 112) = 1.38$, $p = .255$.

Transfer phase The results of the transfer phase are presented in the right panel of Fig. 3. Inspection of the figure reveals that performance on the old items was better in the simultaneous condition ($M = .88$) than in the sequential condition ($M = .81$). Although mean performance on the generalization items ($M = .52$) was well above chance (.33), it was relatively poor overall and did not differ across conditions. To confirm this description, we conducted a 2×2 ANOVA with condition (Simultaneous vs. Sequential) as a between-subjects factor and item type (Old vs. New) as within-subject factor. The analysis revealed no main effect of condition, $F(1, 53) = 2.19$, $p = .145$, but a significant effect of item type, $F(1, 53) = 487.27$, $p < .0001$, $\eta_G^2 = 0.701$, with participants much better at classifying old items than new items. There was also a significant interaction between the two variables, $F(1, 53) = 4.56$, $p = .037$, $\eta_G^2 = 0.021$. Pairwise contrasts showed that whereas there was no difference between the two conditions for new items, $t(53) = 0.19$, $p = .852$,

¹ The pattern of results is the same if all subjects are included in the analyses and none of our conclusions are changed. Because the severe outlier subjects probably failed to understand instructions or had insufficient motivation, we believe it is more sensible to report the data with the severe outliers removed.

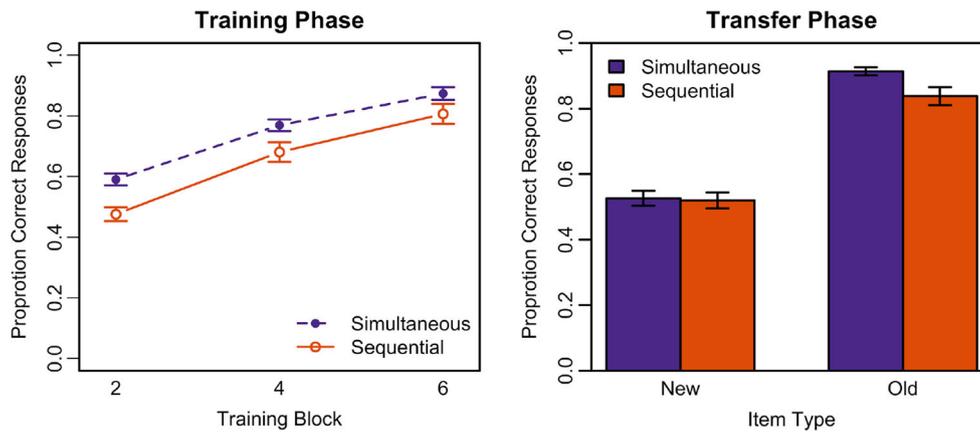


Fig. 3 Results of Experiment 1. The results of the Training phase are presented in the left panel and the results of the Transfer phase are presented in the right panel. Error bars represent standard errors of the mean

participants classified old items significantly better following simultaneous than sequential study, $t(53) = 2.49$, $p = .031$, $d = 0.673$.

Discussion

The learning-phase results and old-item transfer results from Experiment 1 confirm our hypothesis that the use of an organized, simultaneous-presentation display (in combination with sequential testing blocks) can enhance learning of complex natural categories. However, a disappointing result is that the performance benefit did not extend to the generalization items.

To test the possibility that these results were due exclusively to the alternating study-test nature of the simultaneous condition we conducted a follow-up experiment that replicated the results presented here. In the follow-up experiment, we included an additional condition in which the Simultaneous blocks were replaced by blocks that involved a sequential-presentation observational-study procedure (see [Supplementary Materials](#) for details). We found that initial learning of the training items was better in the Simultaneous condition than in either of the pure Sequential conditions. Thus, it is likely that it was the study of the simultaneous displays that played the key role in the results found in this experiment. However, there was again no benefit of simultaneous presentation on generalization to the new transfer items, and overall generalization performance remained relatively poor. Because generalization is an essential feature of categorization, we therefore took steps in Experiment 2 to address this limitation.

Experiment 2

The main goal of Experiment 2 was to take steps that we expected would enhance overall generalization performance and to test whether simultaneous visual displays might lead to

more effective generalization under these modified conditions.

In a study conducted subsequent to our Experiment 1, Nosofsky, Sanders, Meagher, and Douglas ([submitted](#)) collected extensive similarity-scaling data for a set of 360 rock images, including all the images used in our Experiment 1. The similarity-scaling revealed that, by happenstance, several of the rock subtypes included in our Experiment 1 were composed of tokens that were themselves highly dispersed. The intent of the research in the present article was to test categories that were dispersed at the *high* level (i.e., igneous, metamorphic, sedimentary), not at the subtype level (i.e., *rhyolite*, *dolomite*, etc. – see Table 1). Clearly, if new transfer items do not resemble old training items, then it is to be expected that generalization performance will be poor, and techniques that might enhance generalization will not have a chance to show themselves.

Hence, the major change in our experimental design was to modify the stimulus materials. Although we used the same high-level categories and subtypes as in Experiment 1, we deleted individual tokens that we judged to be highly dissimilar to typical items within each subtype category. We replaced these deleted tokens with new ones that we judged to be more similar to the remaining items in the subtype to achieve higher homogeneity within subtypes. Because tokens were assigned randomly to serve as training and transfer items for each subject, this manipulation resulted in both greater homogeneity among members of the training sets, and greater overall similarity of the transfer items to the training items.

Furthermore, to avoid possible ceiling effects on the accuracy measures due to increased within-subtype similarity, we decided to reduce the number of training blocks from six to four. As in Experiment 1, training was again followed by a transfer test involving both the old training instances and new generalization items. A couple of more minor changes to the design are described in the [Methods](#) section.

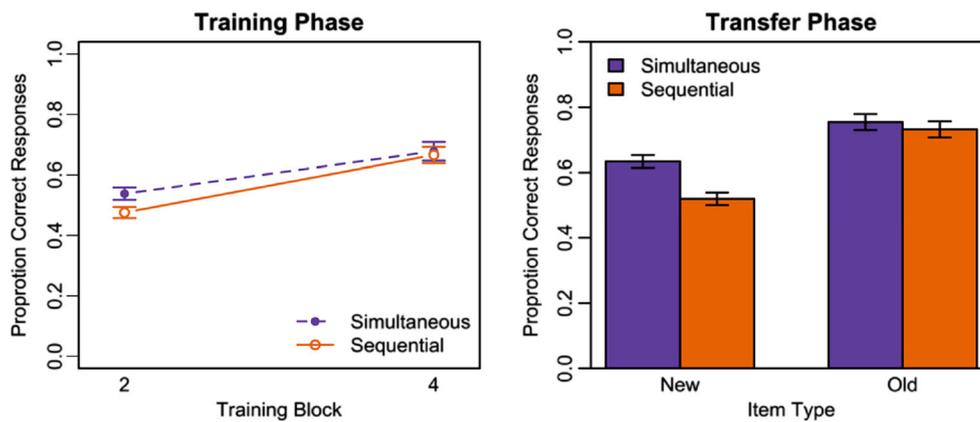


Fig. 4 Results of Experiment 2. The results of the Training phase are presented in the left panel and the results of the Transfer phase are presented in the right panel. Error bars represent standard errors of the mean

Method

Participants Fifty-nine undergraduates from Indiana University participated in partial fulfillment of an introductory psychology course requirement. Participants were randomly assigned to the sequential ($N=32$) or the simultaneous ($N=27$) conditions.

Materials Except for the use of a more homogeneous set of tokens within each subtype (see the [Supplementary Materials](#) for details), all stimulus materials were the same as in Experiment 1. In addition, in the simultaneous-display blocks, the members of each subtype were displayed in an enlarged 2×2 format rather than the 1×4 format of Experiment 1 (see [Supplementary Materials](#) for an illustration). We used this enlarged format so that the size of the images during the training phase matched the size of the images during the test phase.

Procedure The training procedure was the same as in Experiment 1, except we reduced the number of blocks from six to four. (All four training blocks in the sequential condition involved the sequential study-test procedure. The four blocks in the Simultaneous condition were presented in the order simultaneous-sequential-simultaneous-sequential.) In addition, during the transfer phase, we provided corrective feedback on trials in which old training instances were presented, but not on the new-item transfer trials. We followed this procedure with the goal of maintaining subjects' focus and motivation as well as the stability of their memory representations through the duration of the transfer phase.

Results

Pre-processing of outliers Using the same approach described in Experiment 1 (see [Supplementary Materials](#)), we deleted the data of two participants from the transfer-phase

analysis (one from each condition). There were no outliers in the training phase.

Training phase The training-phase results are shown in the left panel of Fig. 4. Although performance was slightly better in the simultaneous condition than in the sequential condition early in training (Block 2), any difference disappeared by Block 4. A 2×2 mixed-model ANOVA revealed a main effect of block of learning, $F(1, 57) = 95.48, p < .0001, \eta_G^2 = .289$. However, there was no main effect of condition, $F(1, 57) = 1.52, p = .223$, and no interaction, $F(1, 57) = 2.14, p = .149$.

Transfer phase The results of the transfer phase are presented in the right panel of Fig. 4. The most important result is that, in contrast to Experiment 1, there is now a robust advantage in correct generalization to new items for the simultaneous group ($M = .63$) compared to the sequential group ($M = .52$). Indeed, for the simultaneous group, overall accuracy on the new items ($M = .63$) is now approaching accuracy on the old training items ($M = .75$). Also in contrast to Experiment 1, for the old test items, there is now only a slight performance advantage for the simultaneous group ($M = .75$) compared to the sequential group ($M = .73$).²

To confirm these observations, we conducted a mixed-model 2×2 ANOVA with condition (Simultaneous vs. Sequential) as a between-subjects factor and item type (Old vs. New) as within-subject factor. The analysis revealed a main effect of condition, $F(1, 55) = 6.56, p = .013, \eta_G^2 = .077$, a main effect of item type, $F(1, 55) = 96.37, p < .0001, \eta_G^2 = .346$, and a significant interaction between the two variables, $F(1, 55) = 6.90$,

² Overall performance on the old items during the transfer phase was worse in Experiment 2 than in Experiment 1. One likely reason is that there were fewer blocks of training in Experiment 2. We discuss a second potential contributing factor in our [General discussion](#).

$p = .011$, $\eta_G^2 = .036$. Pairwise contrasts showed that whereas there was no difference between the two conditions for old items, $t(55) = 0.63$, $p = .533$, participants classified the new items significantly better following simultaneous than sequential study, $t(55) = 4.11$, $p = .0003$, $d = 1.095$.

Discussion

The most important result is that, under the present modified conditions, there is a robust advantage in generalization performance for the simultaneous group compared to the sequential group. Because generalization is an essential feature of categorization, this demonstration speaks strongly to the potential benefits of using organized simultaneous displays as a component of the category-teaching process. Nevertheless, it is of interest that, under the present conditions, performance on the old training items themselves was nearly identical across the simultaneous and sequential conditions. In the [General discussion](#) we advance an account to tie together the complete set of results from Experiments 1 and 2.

General discussion

Summary

Considered across both experiments, the results support the hypothesis that including organized simultaneous displays as part of the teaching method can enhance learning of complex, dispersed natural science categories. However, in Experiment 1 the main benefit of the simultaneous displays was on learning the old training instances with little benefit to generalization performance; in contrast, in Experiment 2 the main benefit was on generalization performance, with little benefit to old-item acquisition.

A unified account

The major difference between the two experiments was that the sets of rock-token pictures used in Experiment 2 had greater within-subtype similarity than did the tokens used in Experiment 1. It follows that in Experiment 2, there are likely more valid generalizations available to be learned than in Experiment 1. For a concrete example, whereas in Experiment 2 virtually all tokens of the subtype quartzite had subtle swirls or veins, such was not the case in Experiment 1, which included a number of quartzite tokens that did not closely resemble the typical instances of this subtype.

We conjecture that the organized simultaneous-display format provides the observer with an enhanced ability to apprehend the overall organization of the rock categories and to determine whether or not valid generalizations are available. If, as in Experiment 2, valid generalizations are available, then

during the learning process the observer focuses attention on the attributes that support those generalizations. In such a case, there will be enhanced generalization to new transfer items, and performance on the new transfer items would tend to approach performance on the old training instances.

But in a case such as Experiment 1, in which valid generalizations are not as readily available, the simultaneous display enables the observer to apprehend this alternative state of affairs. Thus, the observer devotes greater effort towards more idiosyncratic forms of learning, such as attending to idiosyncratic features of individual rocks or engaging in rote memorization. For example, rather than focusing attention on the subtle swirls and veins of the members of the quartzite subtype (which may not even be present in some of the training tokens), the observer may remember instead that a particular token of quartzite had a jagged corner or a salient blemish. These strategic forms of learning and memorization would lead to better performance on the training instances themselves, but not to enhanced generalization, because the idiosyncratic features are less likely to be present on the new transfer items.

Thus, our general account is that the simultaneous-display method provides an advantage across both experiments because it enables learners to use their overview of the whole category space to focus their efforts on the level of information that is most effective to learn that space. In Experiment 1, the most effective learning strategy tends to be to focus on idiosyncratic properties associated with individual training instances, so the greatest advantage is seen on the old training instances themselves. But in Experiment 2, it behooves observers to focus on valid generalizations, so the greatest benefits are on generalization to new items.

Future directions

With future research and continued refinement, our hope is that the type of simultaneous-display method investigated here may eventually be translated for use in real-world geology classrooms. Because our present investigation was motivated around eventual translation, we deliberately used an experimental approach in which multiple factors were expected to lead to the Simultaneous-condition benefit. These multiple factors included: (i) the use of organized (as opposed to disorganized) visual displays in which observers could simultaneously make effective within- and between-category comparisons; (ii) the enabling of self-regulated forms of learning, in which observers could make their own decisions about which parts of the display to focus upon; and iii) the deliberate mixing of the simultaneous-display blocks with testing blocks in order for learners to also receive the well-known benefits of test. Although future research is needed to systematically investigate the influence of each separate factor, we suspect that all these factors make an important contribution to the learning process. We should note as well that there is a remaining

challenge of teaching students to generalize to atypical members of the individual rock subtypes, which we deleted in our Experiment 2 design.

It is interesting to note that our findings have some parallels in a related study concerned with the teaching of statistics principles. In particular, Gurlitt, Dummel, Schuster, and Nückles (2012) recently reported a study involving the use of “structured advance organizers.” Students were presented with structured displays of storylines for statistical analysis. In a well-structured display, each of three columns of the display pertained to cases in which a different statistical test was appropriate to analyzing the data, whereas each of two rows contained storylines with similar surface features. Compared to a control group that used a less structured advance organizer, students provided with the well-structured organizer had better outcomes in understanding the conditions in which to apply the alternative statistical tests. Although the domains of inquiry are quite different, and the statistics problems of Gurlitt et al. (2012) did not require the forms of joint high-level and subtype-level generalization required for learning the present rock categories, it is important that future research identifies common psychological principles that may be operating in our related studies.

Finally, the possibilities for improved teaching of the complex rock categories do not, of course, end with the use of organized simultaneous displays. To take just one example, as noted in our introduction, the scientific categories of igneous, metamorphic, and sedimentary rocks are defined in terms of the manner in which the rocks were originally formed. Although observers do not have direct access to these causal processes, it seems likely that focused instruction on such processes may facilitate students' learning. For example, learning that *granite* is composed primarily of light silicate minerals and that it solidified deep underground may help the student to remember the reason *why* it tends to be light colored and to be composed of coarse grains. Integrating such causal knowledge with effective techniques for the identification of diagnostic perceptual features is likely to lead to even greater facilitation of the learning of scientific classifications.

Acknowledgements This work was supported by NSF Grant 1534014 (EHR Core Research) to Robert Nosofsky and Graduate Training Fellowship SFRH/BD/78083/2011 from the Portuguese Foundation for Science and Technology (FCT) co-sponsored by the European Social Fund to Paulo Carvalho.

References

- Andrews, J. K., Livingston, K. R., & Kurtz, K. J. (2011). Category learning in the context of co-presented items. *Cognitive Processing*, 12(2), 161–175.
- Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, 30(5), 666–677.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- 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.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1147.
- Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393.
- Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, 5(3), 152–158.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38.
- Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*. doi:10.1037/bul0000043
- Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning a cognitive and computational perspective. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(5), 464–481.
- Gurlitt, J., Dummel, S., Schuster, S., & Nückles, M. (2012). Differently structured advance organizers lead to different initial schemata and learning outcomes. *Instructional Science*, 40(2), 351–369.
- Hammer, R., Diesendruck, G., Weinsall, D., & Hochstein, S. (2009). The development of category learning strategies: What makes the difference? *Cognition*, 112(1), 105–119.
- Jee, B. D., Uttal, D., & Gentner, D. (2008). To find fault is easy? The role of comparison in learning a geological structure. In B. Love, K. McRae, & V. Sloutsky (Eds.): *Thirtieth Annual Conference of the Cognitive Science Society*, Washington: DC (p. 1219).
- Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, 10(4), 417–446.
- Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & Cognition*, 43(2), 266–282.
- Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25(4), 431–467.
- Marshak, S. (2013). *Essentials of Geology* (4th Edition). W. W. Norton and Company.
- Nosofsky, R. M., Sanders, C. R., Meagher, B. J., & Douglas, B. J. (submitted). *Toward the Development of a Feature-Space Representation for a Complex Natural Category Domain*. Submitted manuscript.
- Nosofsky, R. M., Sanders, C. R., Gerdman, A., Douglas, B. J., & McDaniel, M. A. (2017). On learning natural science categories that violate the family-resemblance principle. *Psychological Science*, 28, 104–114.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning - Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.
- Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, 32(1-2), 9–31.
- Tarback, E. J., & Lutgens, F. K. (2015). *Earth Science* (14th Edition). Prentice Hall.
- van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, 38(1), 5–13.