Investigating interactions between types of order in categorization

Giulia Mezzadri, Patricia Reynaud-Bouret, Thomas Laloë, Fabien Mathy

Last modified: May 26, 2022

Preprint - Submitted version

Investigating interactions between types of order in categorization

Giulia Mezzadri[1,*], Patricia Reynaud-Bouret[2], Thomas Laloë[2], and Fabien Mathy[3]

[1]Columbia University, Cognition and Decision Lab, New York City, United States
[2]Université Côte d'Azur, Laboratoire J.A. Dieudonné UMR CNRS 7351, Nice, France
[3]Université Côte d'Azur, Bases, Corpus, Langage UMR CNRS 7320, Nice, France

Abstract

This study investigates how different types of presentation order influence category learning and generalization, by simultaneously manipulating within-category (rule-based vs. similarity-based), between-category (blocked vs. interleaved), and across-blocks (constant vs. variable) orders. This research offers a unique and novel approach through both an individual and concurrent analysis of the studied factors (with the investigation of across-blocks manipulations being unprecedented). Both within-category and across-blocks manipulations were found to affect learning, with rule-based and constant orders being more beneficial than similarity-based and variable orders, respectively. Moreover, a significant interaction was found between within-category and across-blocks manipulations. Regarding the classification of novel stimuli, learners in the rule-based and interleaved orders showed generalization patterns that were more consistent with a rule-based strategy, as compared to learners in the similarity-based and blocked orders, respectively.

Keywords: presentation order, category learning and generalization, rule-based vs. similarity-based orders, interleaved vs. blocked study, variable vs. constant across-blocks manipulations

Introduction

What is the best way to memorize lists of vocabulary in another language? Would you study words category by category (for instance, red - blue - yellow - etc., then dog - cat - canary - etc.) to reinforce associations within categories, or would you rather alternate words from different categories (for instance, black - dog - yellow - canary - white - cat - etc.) because you think the two categories can better benefit each other when intertwined? Also, how would you arrange words

* Corresponding author.
E-mail address: gm3026@columbia.edu (G. Mezzadri https://orcid.org/0000-0001-6453-9070).

within a category? Would you first learn words that are phonetically similar (for instance, cat - bat - etc.) or words that are related by a given structure (for instance, warm vs. cold colors) to induce sub-groups, or would you rather study them in random order to facilitate more personal associations from the learner? We believe that these alternative sequences and their combination inevitably produce different outcomes.

A few studies have shown that presentation order influences learning speed and retention in a variety of domains such as memory (Bloom & Shuell, 1981; Farrell, 2008), eyewitness identification (Wells, 2014), serial recall (Miller & Roodenrys, 2012), risk perception (Helsdingen et al., 2011; Kwan et al., 2012), and categorization (Jones & Sieck, 2003; Mack & Palmeri, 2015; Mcdaniel et al., 2013; Sandhofer & Doumas, 2008; Zeithamova & Maddox, 2009; Zotov et al., 2011). In categorization for instance, considerable effort has been directed toward the study of between-category orders (Carvalho & Goldstone, 2014a,b, 2015a, 2021; Goldstone, 1996; Kornell & Bjork, 2008; Kornell et al., 2010; Kost et al., 2015; Noh et al., 2016; Rohrer, 2009, 2012; Sana et al., 2016; Yan et al., 2017; Zulkiply & Burt, 2012; Zulkiply et al., 2012).

More specifically, between-category orders have been thoroughly examined by manipulating interleaving (in which categories are presented alternatively, i.e. a Category-1 member followed by a Category-2 member) vs. blocking (in which members of a single category are presented in a row on successive trials, i.e. a Category-1 member followed by other Category-1 members). In addition to a spacing effect (Carpenter et al., 2012; Carpenter & Mueller, 2013; Cepeda et al., 2008; Hintzman et al., 1975), interleaving stimuli of different categories has been shown to highlight the differences between these stimuli, thus facilitating learning and transfer (Birnbaum et al., 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim et al., 2012; Yan et al., 2017; Zulkiply et al., 2012). However, there has also been evidence in favor of blocking members of a same category (Carpenter & Mueller, 2013; Carvalho & Albuquerque, 2012; Carvalho & Goldstone, 2011; de Zilva & Mitchell, 2012; Rawson et al., 2014).

A lesser number of studies have focused on within-category order effects on category learning (Elio & Anderson, 1981, 1984). Originally explored in word recall (Bower et al., 1969) and old-new recognition tasks (Medin & Bettger, 1994), the manipulation of order within members of a same category has moderately been extended to categorization tasks after the original work of Elio (Corcoran et al., 2011; Mathy & Feldman, 2009, 2016; Stewart et al., 2002). An example of within-category manipulation is the similarity-based order in which stimuli of a same category are arranged in order to maximize the similarity between contiguous examples. This typical manipulation has recently been contrasted with a rule-based order in which stimuli obeying a rule precede the exceptions to the rule. For instance, Mathy and Feldman (Mathy & Feldman, 2016) have found that participants following a "rule plus exceptions" structure show a greater number of generalization patterns consistent with rule-based retrieval than participants in the similarity-based condition. The study of rule-based vs. similarity-based order is particularly relevant since these order manipulations match two extreme ways of learning: an inductive process based on abstraction and an elementary process based on associative mechanisms (Sloman, 1996). The rule-based order is supposed to help participants abstract the logical rule describing the stimuli, while the similarity-based order uses temporal proximity to strengthen the memory traces of contiguous items.

The rationale for the present study is to investigate the effect of presentation orders on memorization and subsequent categorization. While all above-mentioned studies have manipulated a single factor (either between-category or within-category orders), the present study si-

multaneously manipulates between-category orders (blocked vs. interleaved), within-category orders (rule-based vs. similarity-based), and across-blocks orders (constant vs. variable). Effectively, because we used generic types of within-category presentation orders (rule-based vs. similarity-based), the order in which stimuli are presented can still vary once this type of order is chosen. For instance, if we presented a series of faces to be categorized by grouping them by gender or by hair color, we could still apply within-order variations between blocks. A first example of manipulation would consist of varying the order of the faces (within categories) from one block to another. This is what we call across-block variability in our design. Manipulations across-blocks were thought to be particularly important because they have not been addressed by previous studies. For instance, we know that repeating the same sequences using a Hebb-repetition learning procedure increases memorization (French et al., 2011; O'Shea & Clegg, 2006). We therefore hypothesized that maintaining a constant presentation order across blocks could facilitate grouping processes, in comparison to variable orders.

We expected an interaction between within-category orders (rule-based vs. similarity-based) and across-blocks orders (constant vs. variable) because grouping processes should greater benefit the extraction of rules (which in the conception adopted in the present study are more hierarchical by nature than similarity relations). We finally hypothesized a full interaction between our three factors. For instance, we expected a maximal beneficial effect using a rule-based presentation order for which the categories would be blocked and using constant orders across blocks. This combined condition should let participants more easily extract rules and exceptions by focusing on stimuli of one category and by grouping them. On a more theoretical level, the presence or absence of interactions should be informative for current categorization models, which in the future could be adapted to account for differential effects of presentation orders. Any type of model (for instance, rule-based) should be able to account for the conditions that can favor the formation of a particular type of representation (such as a rule). On a more practical level, our findings could have a huge impact on how to organize learning (for instance in the classroom), train experts (for medical image classification), or improve machine learning algorithms.

Because we tested new factors and combinations of factors, we decided to use the widely employed 5-4 category structure from Medin and Schaffer (Medin & Schaffer, 1978) to generate the stimuli and categories, and to study the strategies engaged by participants. A detailed description of the 5-4 category set can be found in Categories. This structure has been analyzed in numerous studies and has influenced research in category learning for more than a quarter century (Blair & Homa, 2003; Cohen & Nosofsky, 2003; Johansen & Kruschke, 2005; Johansen & Palmeri, 2003; Lafond et al., 2007; Lamberts, 2000; Minda & Smith, 2002; Rehder & Hoffman, 2005; Smith & Minda, 2000; Zaki et al., 2003; Nosofsky et al., 1992, 1994; Palmeri & Nosofsky, 1995; Medin & Smith, 1981; Medin et al., 1984; Anderson & Betz, 2001). Moreover, the artificial structure of this category set allows for the presence of stimuli without a category label. The advantage is that on these stimuli different classification strategies lead to distinctive response patterns, allowing us to study the mental representation of the categories (see details in Transfer phase). For these reasons, the 5-4 category set appeared to be a fruitful starting point for our investigation. This study intends to evaluate how different presentation orders impact the speed at which categories are learned, as well as the nature of learning.

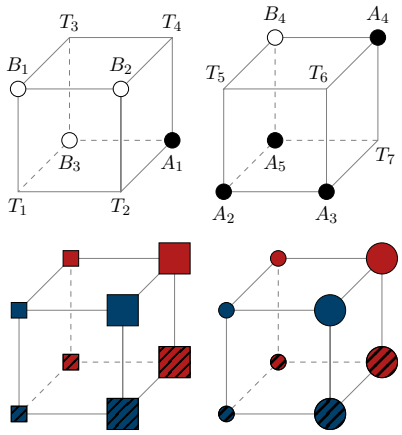

Figure 1

Categories and stimulus items of the categorization task. At the top, the 5-4 category set of Medin and Schaffer (Medin & Schaffer, 1978), represented here in a Hasse Diagram forming a hypercube. Members of category A are represented by black dots, members of category B are represented by white dots, and transfer items are represented by empty vertices. At the bottom, illustration of the stimulus items that varied along four Boolean dimensions (Color, Shape, Size and Filling pattern).

Method

Participants

One hundred and eighty-nine participants contributed to this study. Among them, one hundred and thirty were sophomore or junior students from University Côte d'Azur who received course credits in exchange for their participation. The remaining fifty-nine participants were recruited on campus on a voluntary basis. We used G*Power (Faul et al., 2007) to estimate the power of detecting a small-medium effect size ($f = 0.2$) for the interaction between the three types of order manipulation ($2 \times 2 \times 2 = 8$ between-subject groups) with a three-way ANCOVA model, considering 189 participants, 1 co-variate (i.e., the block number), and $\alpha = 0.05$. The power achieved was 78%. Note that the data-set corresponding to the first one hundred and thirty participants has already been used in (Mezzadri et al., 2022) for testing categorization models. The experimental procedure was approved by the local ethics committee (CERNI #2020-74) of Université Côte d'Azur and the experiment was performed in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants prior to participation.

Categories

Each participant was administrated a single 5-4 category set (Medin & Schaffer, 1978). This structure is composed of 16 stimuli, varying on four different binary-valued dimensions (see Figure 1, on the top). In this category set, 5 stimuli belong to category A, 4 belong to category B, and the remaining 7 are transfer stimuli. These categories are more structured than random (i.e., a clear rule-plus-exceptions pattern emerges) and are linearly separable. The $5 + 4 = 9$ stimuli characterized by a category label were presented in both the learning and transfer phases, whereas the 7 transfer stimuli were presented in the transfer phase exclusively.

Stimuli

Stimuli varied along four Boolean dimensions (Color, Shape, Size, and Filling pattern). The colors were either blue or red; shapes were either square or circle; sizes were either small or big, and filling patterns were either plain or striped. The combination of these options formed $2^4 = 16$ items (see Figure 1, on the bottom). Color distinguished the objects at the front of the hypercube from those at the back, Shape distinguished the objects in the left cube from those in the right cube, Size distinguished the right and left objects within the cubes, and Filling pattern distinguished the objects at the top of the hypercube from those at the bottom. Each dimension was instantiated by the same physical features and the same category structure was applied to these features across participants.

Phases

A learning phase in which participants were instructed to learn the classification of 9 learning stimuli was followed by a transfer phase in which participants were tested upon presentation of 7 novel stimuli (plus the 9 stimuli previously acquired). In the learning phase, both feedback and no-feedback training were used. In particular, two blocks of feedback training (in which the order of the stimuli was manipulated) were followed by one block of no-feedback training (in which stimuli were randomly presented). This pattern was repeated until the end of the learning phase. The use of random blocks with no-feedback allowed us to assess learning with neither order manipulation nor feedback interfering with the measure of performance. The unbalanced ratio of two blocks of feedback training followed by one block of no-feedback training aimed at increasing the influence of our manipulation, with the idea that the random block could still interfere with the learning process. Participants had to correctly classify stimuli in three no-feedback blocks of $5 + 4 = 9$ stimuli (not necessarily consecutive) to complete the learning phase. The choice of three is arbitrary, but appeared to be a good trade-off between maximizing the memorization of the categories and minimizing the duration of the task (a fundamental point considering that the task was conducted online). Participants were given 200 blocks at the most to reach the learning criterion. Once participants met the learning criterion, the transfer phase was initiated. Participants were informed that they successfully completed the learning phase and that the transfer phase was about to start. The transfer phase was composed of 5 blocks of 16 stimuli (the $5+4 = 9$ learning stimuli and the 7 novel stimuli).

Ordering of stimuli

The experiment was characterized by a full factorial design. Three factors were used, each one having two levels: a within-category order manipulation (Rule-based vs. Similarity-based), a between-category order manipulation (Blocked vs. Interleaved), and a manipulation of order across blocks (Variable vs. Constant). The combination of these factors formed eight conditions (e.g., "Rule-based + Interleaved + Constant", etc.). For simplicity purposes, each condition is denoted using the first letter of each type of order. For instance, condition "Rule-based + Interleaved + Constant" is denoted R+I+C. The number of participants assigned to each condition is given in Table 1. As mentioned above, order was only manipulated in the no-feedback blocks of the learning phase.

Within-category order manipulation. In the rule-based order, stimuli were ordered following a "principal rule plus exceptions" structure, meaning that examples obeying the principal rule were presented strictly before the exceptions. The specific "principal rule plus exceptions" structure of our experiment was the following: all striped items belong to category A except for the small red square, while all plain items belong to category B except for the big red circle (see Figure 1). Therefore, items A_1, A_2, A_3, A_5 were strictly presented before item A_4, and items B_1, B_2, B_4 were strictly presented before item B_3. The items belonging to the principal rule (whether belonging to categories A or B) were randomly selected. Presenting stimuli belonging to the dominant rule in a random order was thought to favor an abstraction process, given that other sequences would have increased the risk of temporarily inducing less informative rules, thus delaying learning. Note that instead of using a principal rule based on Filling pattern (plain vs. striped stimuli), we could have used a principal rule based on Shape (circles vs. squares). Indeed, both rules minimize the number of exceptions.

In the similarity-based order, members within a category were presented in a way that maximized the similarity between adjacent learning stimuli. The first stimulus was randomly chosen while subsequent stimuli were (randomly) chosen among those that were the most similar to the immediately previous item. Similarity between two items x and y was computed by counting the number of common features they shared:

$$s_{xy} = \sum_{i=1}^{4} \mathbb{1}_{\{x_i = y_i\}},$$

where x_i and y_i are the feature values of stimuli x and y on dimension i. For instance, the small plain blue circle and the small striped red square have one single feature in common (small), thus their similarity is 1.

Between-category order manipulation. In the blocked study, categories were strictly blocked ($AAAABBBB$ or $BBBBAAAA$), while in the interleaved study categories were strictly alternated ($ABABABAB$). Because of the regularity of both patterns, the introduction of random blocks during learning was necessary. Indeed because of these repetitive patterns, participants could have guessed the correct classification without paying attention to the stimuli. The ratio between blocked (or interleaved) blocks and random blocks is 1:3, as for the feedback/no-feedback blocks. Therefore, a random block with no-feedback always follows two blocks in which categories are blocked (or interleaved) and feedback is provided. Note that in random blocks feedback was never provided, whereas in blocked/interleaved blocks feedback was always provided.

Across-blocks order manipulation. In the constant manipulation across blocks, the same sequence of stimuli (but obeying the constraints of the between- and within-category orders) was presented in all blocks, while in the variable manipulation across blocks the sequence of stimuli varied from one block to another (again, obeying the constraints of the between- and within-category orders).

Procedure

The categorization task was computer-driven and was conducted online. Participants received instructions before the task began. Stimuli were presented one at a time for 3 s on the center of the computer screen. Category A was associated with the up key, while category B was associated with the down key. Participants had to classify the stimulus in one of the two categories (A

Table 1

Number of participants assigned to each of the 8 conditions of the experiment.

	Rule-based		Similarity-based	
	Constant	Variable	Constant	Variable
Blocked	22	25	26	25
Interleaved	25	21	24	21

and B) using these two response keys. Once the key pressed, a feedback indicating the correctness of participants' classification appeared for 1 s at the bottom of the screen (this was the case only in blocks where feedback was provided). If no key was pressed, the text 'too late' appeared for 1 s at the bottom of the screen. In order to encourage learning, the percentage of correct responses in a block was displayed for 1 s at the end of each random block.

Results

Learning phase

Two of our main questions of interest are *i)* whether the speed at which the concept is learned is affected by the types of order, and *ii)* whether there are interactions between these types of order. To answer the first question, we analyzed the time needed by participants to complete the learning phase (in Analysis of the learning times), whereas to answer the second question, we performed an analysis of the interactions using a three-way ANCOVA (in Analysis of the interactions between the types of order). Two additional analyses can be found in Appendix A and B. The first looks at the number of individuals who did not reach the learning criterion, whereas the second analyzes the percentage of correct responses given by participants over the course of the learning phase. Data from 2 participants who pressed no key on more than 15% of the trials were excluded from the analyses.

Analysis of the learning times

Figure 2 shows the average number of blocks which were required for participants to meet the learning criterion as a function of the experimental conditions, taken separately (Figure A) and combined (Figure B). Visually, the rule-based order appears more beneficial than the similarity-based order, and the constant condition appears more beneficial than the variable condition. Note that not all participants were taken into account in Figure 2, since 20 of them did not meet the learning criterion (see Appendix A for details).

To determine which condition led to the fastest learning while accounting for "unsuccessful participants" (i.e., individuals who did not meet the learning criterion), we used two survival analysis techniques: the Kaplan–Meier survival curves and the Cox proportional-hazards model.

Kaplan–Meier survival curves. We used the Kaplan-Meier estimator (Kaplan & Meier, 1958) to estimate the expected duration of time until the successful completion of the learning phase, considering data from participants who did not complete the task as censored. Figure 3 shows the survival probability as a function of block number for each type of order, taken separately (Figure A) and combined (Figure B). The survival probability estimates how likely partici-

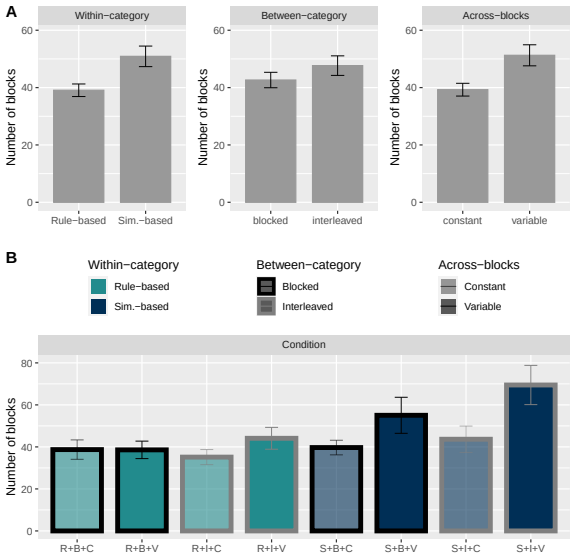

Figure 2

Average number of blocks taken by participants to meet the learning criterion as a function of the experimental conditions, taken separately (Figure A) and combined (Figure B). In Figure B, color distinguishes the within-category order (light blue for rule-based and dark blue for similarity-based), contrast of borders distinguishes the between-category order (black for blocked and gray for interleaved), and opacity of the colors distinguishes the across-blocks order (semi-transparent for constant and opaque for variable). Error bars show $\pm 1SE$*. Note that only the participants that reached the learning criterion were taken into account.*

pants assigned to a given condition are to continue the task (i.e., to not meet the learning criterion). A log-rank test was performed to evaluate the difference between survival curves. The log-rank test was significant at 0.05 level for the within-category and across-blocks orders (for rule-based vs. similarity-based $p = .049$, and for constant vs. variable $p = .0076$). This shows that learning was faster in the rule-based and constant orders as compared to the similarity-based and variable orders, respectively.

Cox proportional-hazards model. Similarly to the Kaplan-Meier estimator, the Cox model (Cox, 1972) allows us to consider failures to complete the task as censored data, avoiding to remove unsuccessful participants. This model is particularly advantageous because of its ability to simultaneously account for multiple variables. Therefore, we use it to simultaneously analyze the influence of the three types of order (within-category, between-category, across-blocks orders) on survival probability. Figure 4A shows the result of the Cox model as a function of our three variables (within-category, between-category, across-blocks orders). The graphs show that the similarity-based order, the interleaved study, and the variable manipulation across-blocks reduced participants' hazard ratio as compared to their respective reference condition (i.e., rule-based order, blocked study, and constant manipulation across-blocks). This means that these types of order were found to reduce participants' speed to meet the learning criterion. However, only the impact of across-blocks manipulations was found significant ($p = .06$ for within-category orders, $p = .132$ for between-category orders, and $p = .009$ for across-blocks orders). Figure 4B shows the

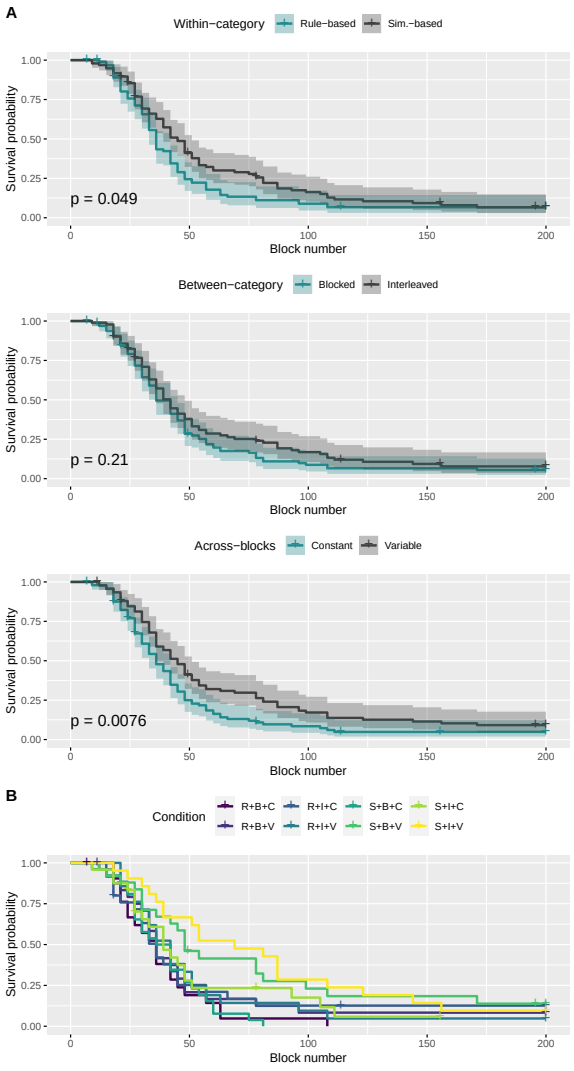

Figure 3

Kaplan-Meier survival curves as a function of block number for each type of order, taken separately (Figure A) and combined (Figure B). Transparent areas represent the 95% confidence intervals. p-values of the log-rang test assessing the difference between survival curves are showed on the bottom-left side of each graph of Figure A.

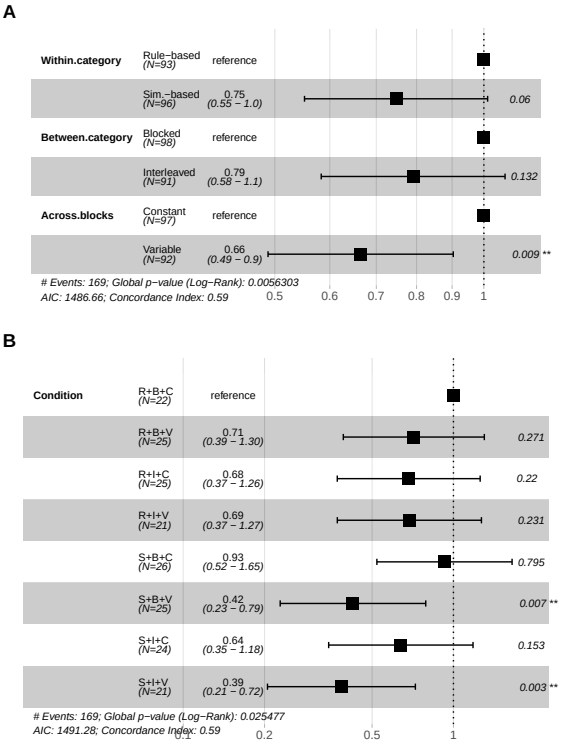

Figure 4

Results of the application of the Cox model as a function of the types of order, taken separately (Figure A) and combined (Figure B). Hazard ratios and their 95% confidence intervals are showed in the middle. Statistical significance of the Wald test is showed on the right.

result of the Cox model as a function of the conditions. Hazard ratio of conditions S+B+V and S+I+V were found to be significantly smaller than the hazard ratio of the reference condition R+B+C ($p = .007$ and $p = .003$, respectively), meaning that participants in conditions S+B+V and S+I+V were statistically slower than participants in condition R+B+C in reaching the learning criterion. We can therefore anticipate on the conclusion of the present study (based on all analyses) that the combination S+V was the most detrimental.

Analysis of the interactions between the types of order

The previous survival analysis techniques allowed us to basically look at main effects. We now aim at searching for interactions between the three types of order manipulation during learning. To do so, we performed a three-way ANCOVA ($2 \times 2 \times 2$ with interactions) with within-category order (rule-based vs. similarity-based), between-category order (blocked vs. interleaved), and across-blocks manipulation (constant vs. variable) as between-subject factors. The number of correct responses per block was the dependent variable and block number was the only co-variate. A probit transformation was applied to the dependent variable in order to meet the assumption of normality. To ensure an equal contribution from each participant, we completed participants' responses until block 72. Since 80% of the participants ended the learning phase before block 72,

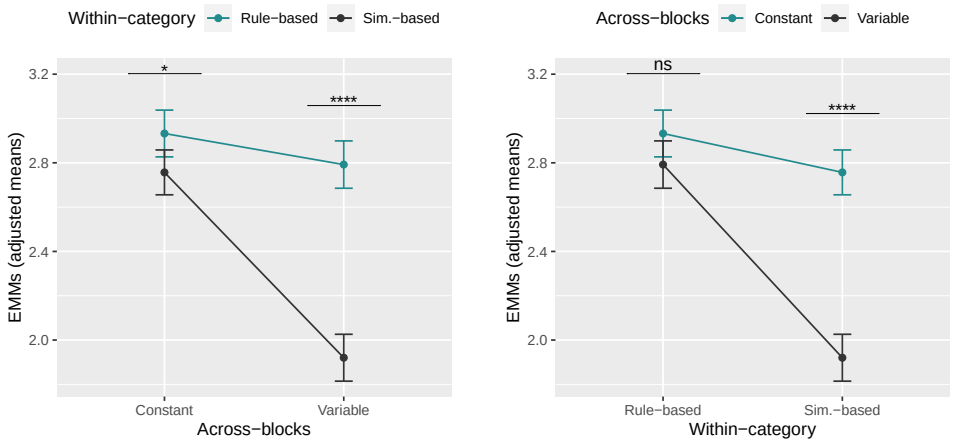

Figure 5

Estimated marginal means (EMMs) for the interaction between within-category and across-blocks orders, with Bonferroni adjustment for multiple testing correction.

this choice allowed us to ensure an equal number of observations for each participant, while limiting the number of observations that were removed or added. Note that we applied the three-way ANCOVA multiple times by varying the block number until which the responses where completed (i.e., block number 48, 51, and 57 corresponding to the 65%, 70%, and 75% quantile). The results were qualitatively the same each time. To complete the data set of participants who ended the experiment before block 72, we iterated their last random block until block 72. We thought it was reasonable to think that participants would have continued to perform in a similar way after ending or abandoning the learning phase. Eighteen observations were found to be multivariate outliers and were excluded from the analysis. Only the interaction between within-category and across-blocks orders was found significant ($F(1, 4461) = 41.52, p < .0001, \eta_p^2 = .009$). To further investigate this interaction, we conducted an analysis of simple main effects, applying Bonferroni adjustment (Dunn, 1961) for multiple testing correction. The simple main effect of within-category order was significant in both the constant ($p < .019$) and variable ($p < .0001$) across-blocks group, while the simple main effect of across-blocks order was only significant in the similarity-based within-category group ($p < .0001$). These effects can be visualized in Figure 5. To complete the analysis, we performed a chi-squared test to assess the difference in performance between the three-way ANCOVA with and without interactions. The difference was significant ($p < .0001$) and the three-way ANCOVA with interactions was the best fitting model.

Transfer phase

Our next aim was to determine whether the types of order affected *i)* performance on learning stimuli and *ii)* generalization patterns on transfer stimuli during the transfer phase. Because we were interested in studying performance and generalization patterns in participants who learned the studied categories, participants who did not meet the learning criterion (amounting to twenty) were excluded from the following analyses. Data from 8 participants who pressed no key on more than 15% of the trials were additionally excluded from the analyses.

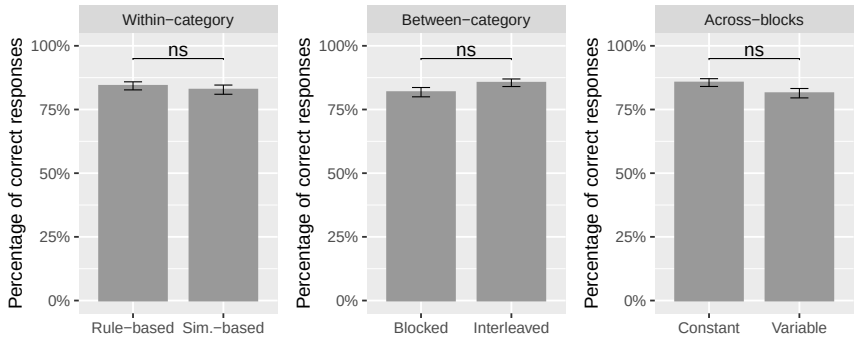

Figure 6

Percentage of correct responses for the learning stimuli presented during transfer, as a function of the types of order (taken separately). Percentage of correct responses was first computed for each participant before being averaged across participants. Error bars show $\pm 1SE$*. Statistical significance of the two-sided Wilcoxon-Mann-Whitney test is shown on the top.*

Analysis of performance on learned stimuli

Figure 6 shows the percentage of correct responses for the learning stimuli presented during transfer, as a function of the types of order (taken separately). The percentage of correct responses was first computed for each participant and then averaged across participants. A two-sided Wilcoxon-Mann-Whitney test was performed to assess the difference in performance between the two conditions within each type of manipulation. None of the tests were found significant ($p = .703$ for rule-based vs. similarity-based, $p = .370$ for blocked vs. interleaved, $p = .057$ for constant vs. variable), showing that performance on learning stimuli was not influenced by the type of order. Although no significant effect was found, one can note that the p-value for constant vs. variable is close to 0.05 and that the direction of the effect (i.e., participants in the constant condition learned better) is coherent with the one that was found in the learning phase (i.e., participants in the constant condition learned faster).

Analysis of generalization patterns on transfer stimuli

Figure 7 shows the average classification of the transfer items over the course of the transfer phase as a function of type of order (taken separately). Quantity $p(A)$ is the observed proportion that each transfer item was classified into category A during transfer. To determine whether participants in different conditions applied different strategies for the classification of novel stimuli, we computed the distance of the observed generalization patterns to four specific strategies (distances were computed using the $L1$ metric and were normalized).

The following strategies were considered: a rule-based strategy that uses Filling pattern (plain vs. striped stimuli) as main rule, a rule-based strategy that uses Shape (circles vs. squares) as main rule, a similarity-based strategy, and a random strategy. Participants adopting a rule-based strategy would classify new stimuli on the basis of the main rule (Filling pattern or Shape, depending on the chosen main rule), whereas participants adopting a similarity-based strategy would classify new stimuli on the basis of their similarity to the closest stored items. In the random strategy, novel stimuli would be randomly classified (50% of chance to classify them into category

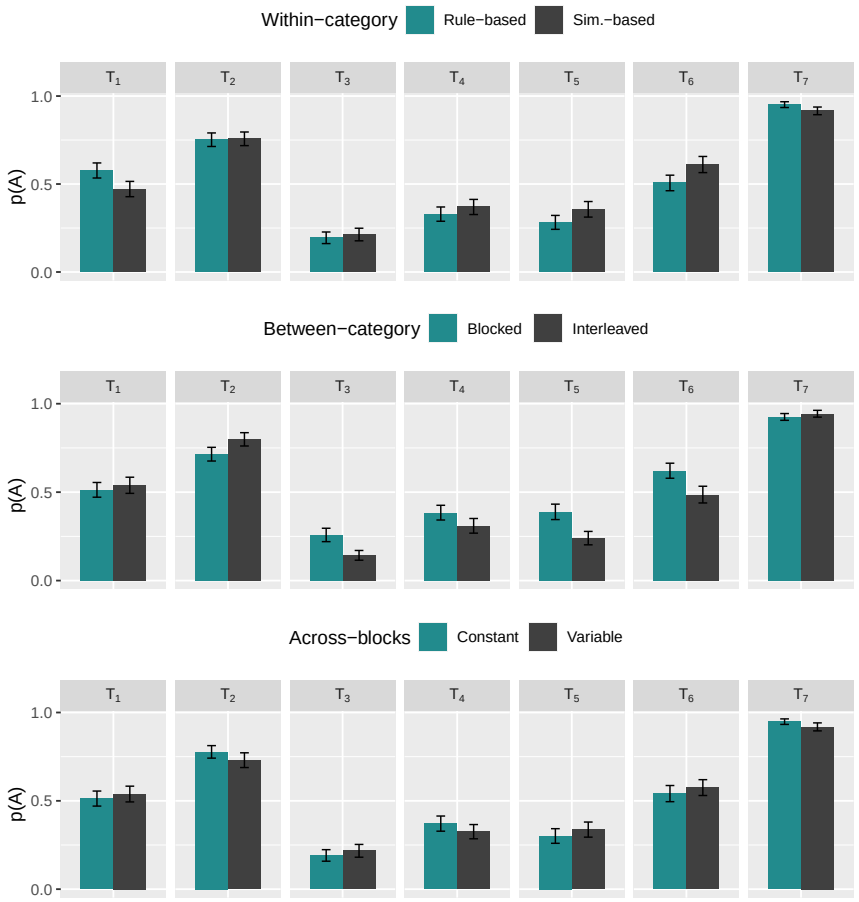

Figure 7

Average classification of the transfer items (T_1, T_2, T_3, T_4, T_5, T_6, T_7) during the transfer phase (amounting to five blocks) as a function of types of order (taken separately). Quantity $p(A)$ is the observed proportion that each of the stimuli labeled under the abscissa was classified into category A during the transfer phase. Quantity $p(A)$ was first computed for each participants before being averaged across participants. Average classification of the nine learning items is not included. Error bars show $\pm 1SE$.

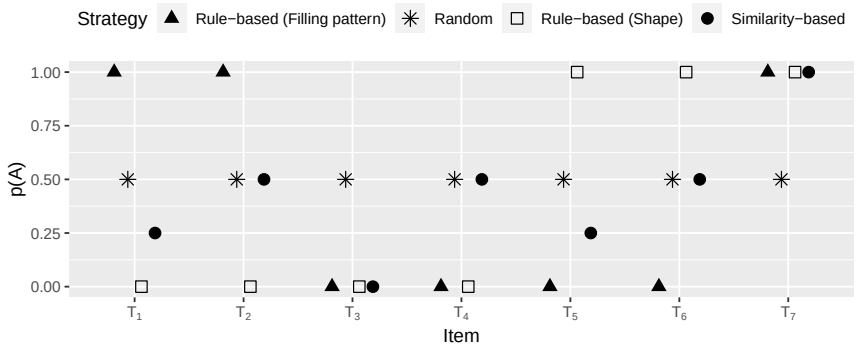

Figure 8

Putative classification of the transfer items (T_1, T_2, T_3, T_4, T_5, T_6, T_7) as a function of the applied strategy (rule-based strategy using Filling pattern, rule-based strategy using Shape, similarity-based strategy, and random strategy). Quantity $p(A)$ is the putative probability for a chosen strategy to classify into category A each of the stimuli labeled under the abscissa.

A). The rule-based strategy that uses Shape as main rule was included because Shape (as Filling pattern) allows participants to minimize the number of exceptions when used as the diagnostic dimension (see Ordering of stimuli). Putative classification of the transfer stimuli for each of the above-mentioned strategies is shown in Figure 8.

Figure 9A shows the average distance of the observed generalization patterns to the above-mentioned strategies. The closest strategies to the observed generalization patterns are the similarity-based strategy and the Filling pattern rule-based strategy, followed by the random strategy, and finally by the Shape rule-based strategy. A two-sided Wilcoxon-Mann-Whitney test was performed to assess the difference in distribution between the different strategies. We found a significant difference between Shape rule-based strategy and random strategy ($p < .0001$), and between random strategy and Filling rule-based strategy ($p < .0001$), but not between Filling rule-based strategy and similarity-based strategy ($p = .458$). Since the Shape rule-based strategy and the random strategy were the farthest to the observed patterns, they were excluded from the following analyses.

Figure 9B-C-D shows the average distance of participants' generalization patterns to two specific strategies (the Filling pattern rule-based strategy and the similarity-based strategy), as a function of the type of order within each manipulation (within-category in Figure B, between-category in Figure C, and across-blocks in Figure D). While distances to the similarity-based strategy were similar across types of order within the same manipulation, distances to the Filling pattern rule-based strategy largely varied. More specifically, generalization patterns of participants in the rule-based order were significantly closer to the Filling pattern rule-based strategy than participant in the similarity-based strategy ($p = .045$). The same applies to participants in the interleaved study as compared to those in the blocked study ($p = .004$). An alternative analysis leading to the same conclusion can be found in Appendix C.

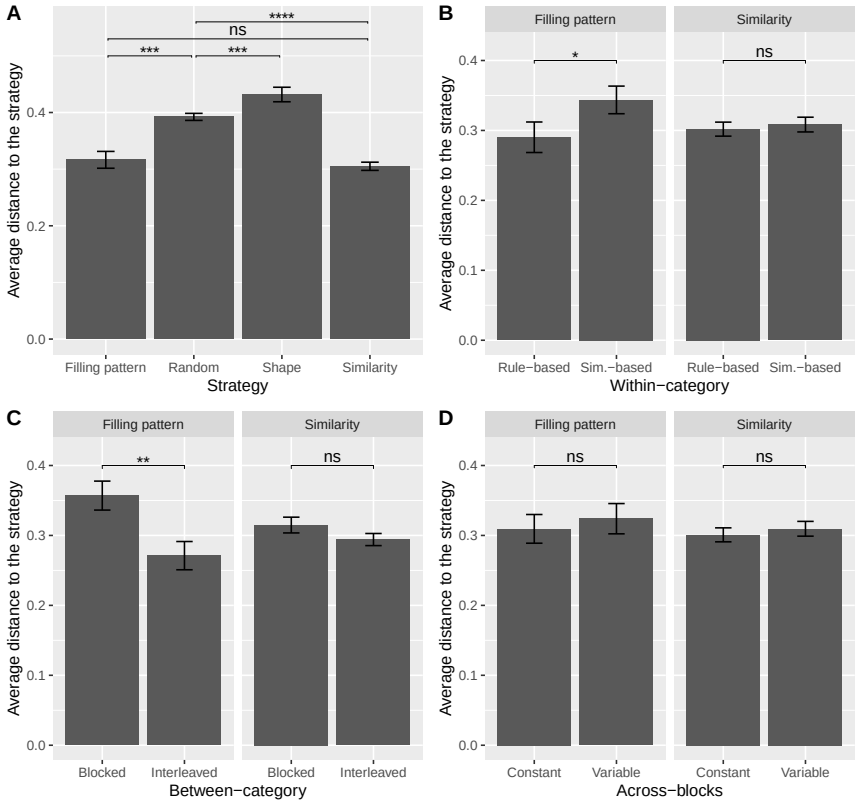

Figure 9

Average distance of participants' generalization patterns on novel stimuli to specific strategies (rule-based strategy using Filling pattern, rule-based strategy using Shape, similarity-based strategy, and random strategy), for all participants (Figure A) and for each type of order within the same manipulation (Figure B-C-D). Distances were first computed for each participant before being averaged. The $L1$ norm was used and distances were normalized prior to averaging. Asterisks show the significance of the two-sided Wilcoxon-Mann-Whitney test.

Discussion

Previous studies on category learning have shown that the sequence in which stimuli are encountered can profoundly influence learning speed and category formation (Carvalho & Goldstone, 2021; Kang & Pashler, 2012; Mathy & Feldman, 2016). However, the totality of these studies have only focused on a subset of types of presentation orders (either within-category or between-category manipulations), separately, ignoring potential interactions between different types of manipulations. Here, we manipulated within-category, between-category, and across-blocks order manipulations within a single task to study their separate and combined impact on learning speed and generalization patterns. The main question of interest is whether these types of order manipulation *i)* influence how fast categories are learned, *ii)* interact with one another during learning, and *iii)* affect classification of novel stimuli. We here used a single category structure, namely the '5-4', to which we applied our combined manipulations of presentations orders. Our result most probably reflects the nature of this specific category structure. We have no doubt that

other effects could apply to other structures. The question here is simply whether presentation orders can potentially modify learning and category representations. If yes, future research should attempt to generalize our observations to other cases.

To address the first point, we analyzed how long participants took to reach the learning criterion as a function of the conditions within each order manipulation. Both the Kaplan-Meier and Cox survival analyses showed that the pace at which categories are learned is influenced by both within-category and across-blocks orders (with the across-blocks manipulation being the main predictor of participants' learning speed). More specifically, the rule-based (within-category) and constant (across-blocks) order manipulations were found to lead to faster learning than the similarity-based (within-category) and variable (across-blocks) order manipulations, respectively. Moreover, participants in the condition combining similarity-based within-category order and variable across-blocks order have been found to be the slowest at learning the categories.

However, these findings alone do not allow us to determine which condition is the most beneficial. A further analysis of performance on learned stimuli during transfer was more adequate to reach this goal. Indeed, conditions that lead learners to more rapidly reach the learning criterion are not always the ones that lead to better learning (Soderstrom & Bjork, 2015). No significant effect of order manipulation was found on participants' performance on learned stimuli during transfer. Moreover, the only effect close to significant that showed better learning in the constant condition (as compared to the variable one) went in the same direction as the effect found in the learning phase (i.e., faster learning in the constant condition). We can thus conclude that the rule-based and constant conditions are more beneficial to learning, since performance (on learned stimuli) during transfer was statistically the same for every type of order, but participants in these conditions were exposed to a fewer number of trials.

Following the interpretation of Mathy and Feldman (Mathy & Feldman, 2009), the benefit of the rule-based order over the similarity-based order might be imputed to the illusory proximity of stimuli that the similarity-based order might induce in participants' mind. The similarity-based order promotes a structure that is not fully informative about the categories, which might temporarily mislead participants. This is especially true for the 5-4 category set that is often considered 'rule-based' in itself because of its use of discrete features. By contrast, the rule-based order first presents the most informative information, which could yield superior learning. The superiority of the constant across-blocks presentation over the variable one might be attributed to the limited amount of information carried by the constant order. Limiting the variability of the sequences might have helped participants focus on diagnostic information, enhancing the probability to either induct the simplest rule or memorize the category membership of the items. Another explanation is that memory can benefit from the repetition of the same sequences (French et al., 2011; O'Shea & Clegg, 2006), in particular to group items. A grouping process could have benefited the formation of rules.

To determine whether different types of order interact with one another during learning, we performed statistical analyses with and without interactions. Data were best explained by the model with interactions, that lead to one significant interaction between within-category and across-blocks orders. In particular, the similarity-based order (in comparison to the rule-based order) was detrimental for participants in both the constant and variable conditions (with the variable condition being the most detrimental), whereas the variable condition (in comparison to the constant condition) was only detrimental for participants in the similarity-based order. Therefore, the condition combining similarity-based within-category order with variable across-blocks order

was the most detrimental for learning.

Our last analysis aimed at determining the influence of order manipulation on the classification of novel stimuli by analyzing how close are the observed patterns to some main strategies. The rule-based strategy that uses Filling pattern as the main rule as well as the similarity-based strategy were found to be the closest to participants' generalization patterns. Moreover, distance to the Filling pattern rule-based strategy was significantly higher for participants in the similarity-based order (as compared to those in the rule-based order) and for participants in the blocked order (as compared to those in the interleaved order). By contrast, across-blocks manipulations were not found to alter category transfer. These results show that both within-category and between-category orders affect how learning is transferred to novel stimuli, with the Filling pattern rule-based strategy being preferred more often in the rule-based and interleaved conditions than in the similarity-based and blocked conditions, respectively.

This difference in preference between rule-based and similarity-based learners might be attributed to the logic upon which the rule-based order is grounded. Presenting items following a "main rule (Filling pattern) plus exceptions" structure might have facilitated participants to infer the simplest rule, encouraging them to classify new items using the same inferred strategy. Conversely, the difference in preference between blocked and interleaved learners might find an explanation in the Sequential Attention Theory (Carvalho & Goldstone, 2015b). Alternating stimuli from different categories (interleaved study) leads to an attentional focus on properties that discriminate the categories, which might have promoted a rule-based transfer of the knowledge. By contrast, when stimuli from the same category are presented sequentially (blocked study), the encoding of the similarities among items of the same category is stronger, which might have promoted the use of a similarity-based strategy. In addition, the structure of the 5-4 category set (in particular the fact that the within-category similarity exceeds the between-category similarity) might also have played a role.

An additional contribution of the present study is the promotion of underemployed statistical tools. A common practice in psychology is to remove participants who did not fulfill the objective of the task (Mathy & Feldman, 2016; Meagher et al., 2017, 2018). Nevertheless, unsuccessful participants can carry useful information. In the present study, we made use of two survival analysis techniques (the Kaplan-Meier survival curves and the Cox model) that allow us to account for individuals who did not complete the task. We advise the use of similar statistical tools when the conditions allow them.

Study limitations

Although Medin and Shaffer's 5-4 category set has multiple benefits (see Introduction), it has the disadvantage of presenting a clear "rule plus exceptions" structure and of being distant from real-world categories. Moreover, the absence of transfer stimuli with a category label did not allow us to generalize the benefit of the rule-based and constant orders to novel stimuli. However, because it has been thoroughly scrutinized in the past, this structure seemed the best option as first investigation of new experimental factors. Another limitation of this category structure is that it allows the co-existence of two different main rules that minimizes the number of exceptions (Filling pattern and Shape). Stimuli in the rule-based order were ordered following the "Filling pattern rule plus exceptions" structure. However, the dimension Shape could have equivalently been used instead of the dimension Filling pattern. In the same spirit, dimensions could have been instantiated by different features. For instance, Color could have distinguished the right and left

objects within the cubes, Shape the objects at the front of the hypercube from those at the back, Size the objects at the top of the hypercube from those at the bottom, and Filling pattern the objects in the left cube from those in the right cube. We felt that multiplying our sample to consider all above-mentioned variations would have been too costly.

Acknowledgements

We are thankful to all of the volunteers who took part to our experiment. This work was supported by the French government, through the UCAJedi and 3IA Côte d'Azur Investissements d'Avenir managed by the National Research Agency (ANR-19-P3IA-0002 and ANR-15-IDEX-01), directed by the National Research Agency with the ANR project ChaMaNe (ANR-19-CE40-0024-02) and by the interdisciplinary Institute for Modeling in Neuroscience and Cognition (NeuroMod) of the Université Côte d'Azur.

Author contributions

Study concept and study design were developed by G.M. with contributions from F.M., P.R.-B. and T.L. The categorization task was implemented in PsychoPy by G.M. Testing, data collection, and data analysis were performed by G.M. The main manuscript was wrote by G.M. and critical revisions were provided by F.M., P.R.-B. and T.L. All authors approved the final version of the manuscript for submission.

Data availability

Data generated during the current study are available in Open Science Framework at https://osf.io/w29ts/?view_only=c28b965cc9a74c54b56d7adb87417ff1.

Additional information

Competing interests. The authors declare no conflict of interest with respect to their authorship or the publication of this article.

References

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, *8*(4), 629–647.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Birnbaum, M., Kornell, N., Bjork, E., & Bjork, R. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*, 392–402.

Blair, M., & Homa, D. (2003). As easy to memorize as they are to classify: The 5–4 categories and the category advantage. *Memory & Cognition*, *31*(8), 1293–1301.

Bloom, K., & Shuell, T. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, *74*, 245–248.

Bower, G., Clark, M., Lesgold, A., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343.

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

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*, 369–378.

Carvalho, P. F., & Albuquerque, P. B. (2012). Memory encoding of stimulus features in human perceptual learning. *Journal of Cognitive Psychology*, *24*, 654–664.

Carvalho, P. F., & Goldstone, R. L. (2011). Sequential similarity and comparison effects in category learning. *Proceedings of the 33rd annual conference of the Cognitive Science Society*, *33*, 2977–2982.

Carvalho, P. F., & Goldstone, R. L. (2014a). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology*, *5*, 936.

Carvalho, P. F., & Goldstone, R. L. (2014b). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*, 481–495.

Carvalho, P. F., & Goldstone, R. L. (2015a). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*, 281–288.

Carvalho, P. F., & Goldstone, R. L. (2015b). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*, 505.

Carvalho, P. F., & Goldstone, R. L. (2021). The most efficient sequence of study depends on the type of test. *Applied Cognitive Psychology*, *35*, 82–97.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: a temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102.

Cohen, A. L., & Nosofsky, R. M. (2003). An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, *47*(2), 150–165.

Corcoran, K., Epstude, K., Damisch, L., & Mussweiler, T. (2011). Fast similarities: Efficiency advantages of similarity-focused comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1280–1286.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*, 187–220.

de Zilva, D., & Mitchell, C. (2012). Effects of exposure on discrimination of similar stimuli and on memory for their unique and common features. *Quarterly Journal of Experimental Psychology*, *65*, 1123–1138.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American statistical association*, *56*(293), 52–64.

Elio, R., & Anderson, J. (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.

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 128 – 145.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*, 614–636.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*(608-628).

Helsdingen, A., Gog, T., & Van Merrienboer, J. J. G. (2011). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*, *21*, 126–136.

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). What causes the spacing effect? some effects of repetition, duration, and spacing on memory for pictures. *Memory & Cognition*, *3*, 287–294.

Johansen, M., & Kruschke, J. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1433–1458.

Johansen, M., & Palmeri, T. (2003). Are there representational shifts in category learning? *Cognitive Psychology*, *45*, 482–553.

Jones, M., & Sieck, W. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 626–640.

Kang, S., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.

Kornell, N., & Bjork, R. (2008). Learning concepts and categories: is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592.

Kornell, N., Castel, A., Eich, T., & Bjork, R. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*, 498–503.

Kost, A. S., Carvalho, P. F., & Goldstone, R. L. (2015). Can you repeat that? the effect of item repetition on interleaved and blocked study. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, (pp. 1189–1194).

Kwan, V. S. Y., Wojcik, S. P., Miron-Shatz, T., Votruba, A. M., & Olivola, C. Y. (2012). Effects of symptom presentation order on perceived disease risk. *Psychological Science*, *23*, 381–385.

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology*, *51*, 57–74.

Lamberts, K. (2000). Information accumulation theory of categorization. *Psychological Review*, *107*, 227–260.

Mack, M., & Palmeri, T. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*, 551–569.

Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, *16*, 1050–1057.

Mathy, F., & Feldman, J. (2016). The influence of presentation order on category transfer. *Experimental Psychology*, *63*, 59–69.

Mcdaniel, M., Fadler, C., & Pashler, H. (2013). Effects of spaced versus massed raining in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1417–1432.

Meagher, B., Carvalho, P., Goldstone, R., & Nosofsky, R. (2017). Organized simultaneous displays facilitate learning of complex natural science categories. *Psychonomic Bulletin & Review*, *24*, 1987–1994.

Meagher, B., Cataldo, K., Douglas, B., Mcdaniel, M., & Nosofsky, R. (2018). Training of rock classifications: The use of computer images versus physical rock samples. *Journal of Geoscience Education*, *66*, 1–10.

Medin, D., & Bettger, J. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, *1*, 250–254.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 333.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(4), 241.

Mezzadri, G., Reynaud-Bouret, P., Laloë, T., & Mathy, F. (2022). An order-dependent transfer model in categorization. *Journal of Mathematical Psychology*, *107*, 102634.

Miller, L. M., & Roodenrys, S. (2012). Serial recall, word frequency, and mixed lists: the influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1731–1740.

Minda, J., & Smith, J. (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.

Noh, S., Yan, V., Bjork, R., & Maddox, W. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, *155*, 23–29.

Nosofsky, R., Gluck, M., Palmeri, T., Mckinley, S., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of shepard, hovland, and jenkins (1961). *Memory & Cognition*, *22*, 352–369.

Nosofsky, R., Kruschke, J., & McKinley, S. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 211–233.

O'Shea, G., & Clegg, B. A. (2006). Stimulus and response chunking in the hebb digits task. *Psychological Research*, *70*, 180–192.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 548.

Rawson, K., Thomas, R., & Jacoby, L. (2014). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*, 483–504.

Rehder, B., & Hoffman, A. (2005). Thirty-something categorization results explained: Attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 811–829.

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, *40*, 4–17.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*, 355–367.

Sana, F., Yan, V., & Kim, J. (2016). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*, 84–98.

Sandhofer, C., & Doumas, L. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development - J COGN DEV*, *9*, 194–221.

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Smith, J., & Minda, J. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 3–27.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199.

Stewart, N., Brown, G., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 3–11.

Wahlheim, C., Finn, B., & Jacoby, L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: Evidence for variability neglect. *Memory & Cognition*, *40*, 703–716.

Wells, G. L. (2014). Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science*, *23*, 11–16.

Yan, V., Soderstrom, N., Seneviratna, G., Bjotk, E., & Bjork, R. (2017). How should exemplars be sequenced in inductive learning? empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied*, *23*, 403–416.

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1160–1173.

Zeithamova, D., & Maddox, W. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 731–741.

Zotov, V., Jones, M., & Mewhort, D. (2011). Contrast and assimilation in categorization and exemplar production. *Attention, perception & psychophysics*, *73*, 621–639.

Zulkiply, N., & Burt, J. (2012). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16–27.

Zulkiply, N., Mclean, J., Burt, J., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215–221.

Appendix A
Analysis of the number of unsuccessful participants

The term "unsuccessful participants" refers to those individuals who did not meet the learning criterion, either because they dropped out or because they exceeded the maximal amount of blocks allowed to the experiment (200 learning blocks). Table A1 shows the number of successful and unsuccessful participants per types of order, taken both separately and combined. A Fisher's exact test of independence at 0.05 level was run to determine whether the number of individuals who did not reach the learning criterion was related to the type of order. The Fisher's exact test was preferred to the chi-square test because of its accuracy with small samples. None of the tests were significant ($p = .461$ for rule-based vs. similarity-based, $p = .325$ for blocked vs. interleaved, $p = 0.624$ for constant vs. variable, and $p = .088$ for the omnibus test run on the table showing all conditions). We conclude that the type of order in which stimuli were encountered did not alter participants' chance of reaching the learning criterion.

Appendix B
Analysis of the percentage of correct responses during learning

Because we were interested in analyzing the learning curves of participants who learned the studied categories, unsuccessful participants (amounting to twenty) were excluded from the analysis. Figure B1 shows the average percentage of correct responses among participants within a same type of order taken separately (Figure A) and combined (Figure B), as a function of block number

Table A1

Number of successful and unsuccessful participants split per type of order, taken both separately and combined.

Type of order	Successful	Unsuccessful
Within-category orders		
Rule-based	84	7
Similarity-based	85	11
Between-category orders		
Blocked	89	7
Interleaved	80	11
Across-blocks orders		
Constant	88	8
Variable	81	10
Conditions		
R + B + C	21	0
R + B + V	22	2
R + I + C	21	4
R + I + V	20	1
S + B + C	26	0
S + B + V	20	5
S + I + C	20	4
S + I + V	19	2

over the course of the learning phase. Only participants' performance during random blocks is represented in the graph. Because it was reasonable to think that successful participants would have continued to correctly classify stimuli after reaching the learning criterion, their responses were completed with the highest performance (i.e., 100% of correct responses) until block number 72. Because 80% of the participants ended the learning phase before block 72, this choice appeared to be a good trade-off between minimizing the number of observations that were both removed and added. In Figure B1A, learning was more efficient in the rule-based, blocked, and constant orders than in the similarity-based, interleaved, and variable orders, respectively. To assess the difference between the learning curves, we performed a two-sided Wilcoxon-Mann-Whitney test at each block with Benjamini-Hochberg adjustment (Benjamini & Hochberg, 1995) for multiple testing corrections. In the within-category and across-blocks manipulations, half of the test were significant (11/24 rejected tests for the within-category and 12/24 rejected tests for the across-blocks), whereas in the between-category order none of the tests were significant. We can conclude that in half of the random trials learning in the rule-based and constant conditions was more effective than the similarity-based and variable conditions (respectively).

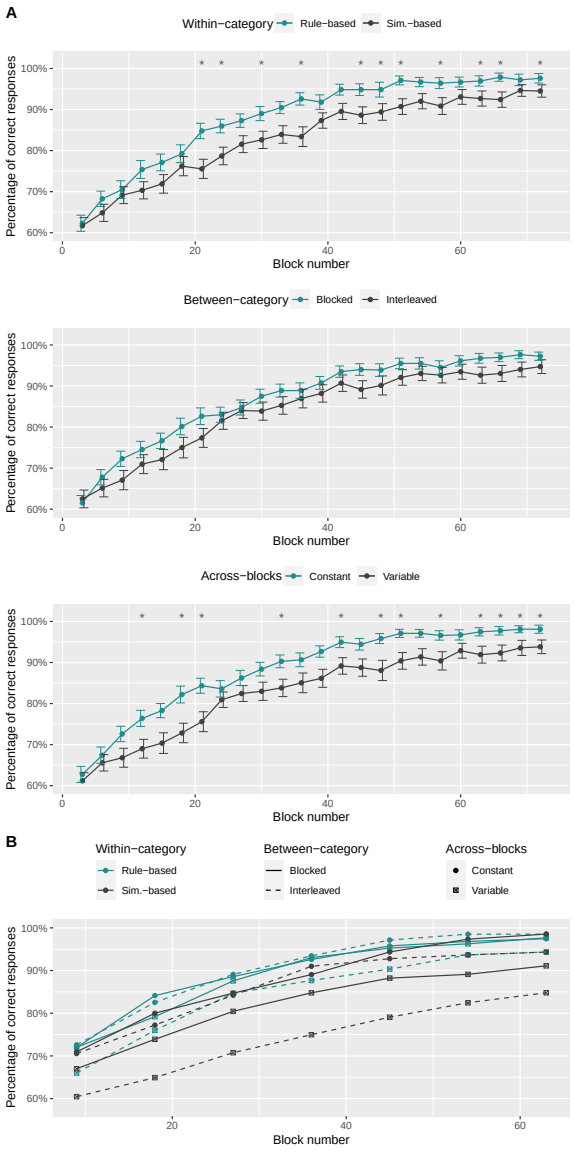

Figure B1

Average percentage of correct responses among participants within a same type of order taken separately (Figure A) and combined (Figure B), as a function of block number over the course of the learning phase. In Figure A, only performance across random blocks are plotted. Asterisks indicate the blocks on which the Wilcoxon-Mann-Whitney test with the Benjamini-Hochberg correction was rejected. Error bars show $\pm 1SE$*. In Figure B, color distinguishes the within-category order (blue for rule-based and gray for similarity-based), line-type distinguishes the between-category order (solid line for blocked and dashed line for interleaved), and shape distinguishes the across-blocks order (dots for constant and crossed squares for variable). To increase smoothness of the curves, performance at each random block are obtained by averaging performance from the two preceding random blocks, the current random block, and the two following random blocks. To make the plot more readable, only performance every three random blocks are plotted.*

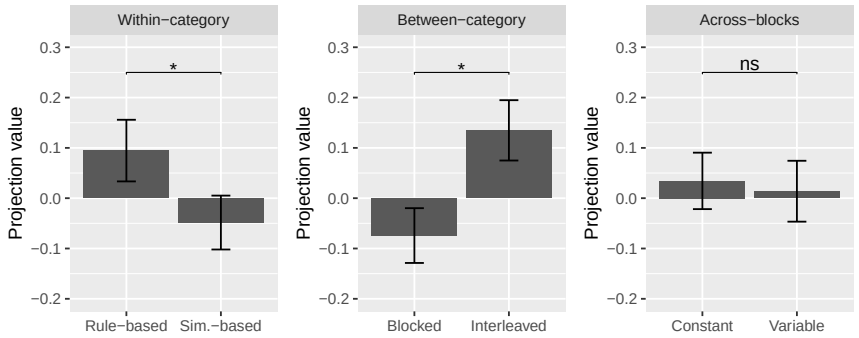

Figure C1

Projection value of the observed generalization patterns, as a function of different conditions within each order manipulation. Generalization patterns were projected on the one-dimensional space generated by vector $v = (1, 1, 0, -1, -1, -1, 0)$*, which distinguishes the Filling pattern rule-based strategy from the similarity-based strategy. Error bars show* $\pm 1SE$*. Asterisks show the significance of the two-sided Wilcoxon-Mann-Whitney test.*

Appendix C
Alternative analysis of generalization patterns on transfer stimuli

The aim of this analysis is to determine whether participants in different conditions classify new items using different strategies. We focus on two main strategies: a rule-based strategy that uses Filling pattern as main rule and a similarity-based strategy (see Figure 8). A learner adopting the Filling pattern rule-based strategy would more often classify items T_1 and T_2 into category A and items T_4, T_5 and T_6 into category B, than a learner adopting the similarity-based strategy. Therefore, these two strategies can be distinguished by projecting generalization patterns on the one-dimensional space generated by vector $v = (1, 1, 0, -1, -1, -1, 0)$. The more the projections are on the right side of this one-dimensional space, the more the Filling pattern rule-based strategy was used. Figure C1 shows the projection value of participants' generalization patterns on the one-dimensional space generated by vector v as a function of the conditions within each order manipulation. A two-sided Wilcoxon-Mann-Whitney test was conducted to assess the difference between projection values of participants in different conditions. Projections from participants in the rule-based order were significantly higher than those from participants in the similarity-based order ($p = .043$). Similarly, projections from participants in the interleaved order were significantly higher than those from participants in the blocked order ($p = .010$). No significant difference was found between projections from participants with different across-blocks orders ($p = .919$). We conclude that participants in the rule-based and interleaved orders showed generalization patterns that where more consistent with the Filling pattern rule-based strategy than participants in the similarity-based and blocked orders, respectively.