Investigating interactions between types of order in categorization

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

Investigating interactions between types of order in categorization

Giulia Mezzadri

Université Côte d'Azur

Patricia Reynaud-Bouret

Université Côte d'Azur

Thomas Laloë

Université Côte d'Azur

Fabien Mathy

Université Côte d'Azur

Author Note

Giulia Mezzadri https://orcid.org/0000-0001-6453-9070

Correspondence concerning this article should be addressed to Giulia Mezzadri, Laboratoire J.A. Dieudonné UMR CNRS 7351, Université Côte d'Azur, Parc Valrose 06108 Nice Cedex 2, France. E-mail: giulia.mezzadri@univ-cotedazur.fr

All data, analysis code, and study materials are available in the Open Science Framework at https://osf.io/w29ts/?view_only=c28b965cc9a74c54b56d7adb87417ff1.

Abstract

This study investigates how different types of presentation orders influence category learning and generalization. We used a full factorial design (within-category, between-category, and across-blocks manipulations), each factor having two levels (rule-based vs. similarity-based, blocked vs. interleaved, and constant vs. variable orders, respectively). This research offers a unique and novel approach through both an individual and concurrent analysis of the studied factors. Moreover, the investigation of across-blocks manipulations is unprecedented. Using survival analysis techniques as well as non-parametric tests, we found that learning was impacted by both within-category and across-blocks manipulations (with across-blocks manipulation being the main predictor). More specifically, the rule-based and constant orders were found more beneficial than the similarity-based and variable orders, respectively. Also, within-category and between-category orders were found to affect generalization patterns. More precisely, participants in the rule-based order more often showed generalization patterns consistent with a rule-based strategy than participants in the similarity-based order. Likewise, participants in the interleaved order more often showed generalization patterns consistent with a rule-based strategy than participants in the blocked order. We conclude that combining a similarity-based order with a variable across-blocks manipulation delays learning, whereas combining a rule-based order with a constant across-blocks manipulation benefits learning. Furthermore, combining a rule-based order with an interleaved order favors the use of a rule-based strategy during category generalization.

Keywords: Presentation order, Category learning, Category generalization, Rule-based vs. similarity-based orders, Interleaved vs. blocked study, Variable vs. constant across-blocks manipulations

Investigating interactions between types of order in categorization

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 - green - yellow - etc.) or would you alternate words from different categories (for instance, red - dog - hoody - blue - cat - shoes - etc.)? Also, how would you arrange words within a category? Would you first learn words that are phonetically similar (for instance, my - buy - cry - high - etc.), or words that are related by a given structure (for instance, warm vs. cold colors), or would you rather study them in random order? 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, 2014b, 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 (S. Carpenter & Mueller, 2013; S. K. Carpenter et al., 2012; 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 (S. 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; Elio & Anderson, 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 (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.

All the above-mentioned studies have manipulated a single factor (either between-category or within-category order). We believe that studying the interactions between these factors is crucial. For instance, participants in the rule-based order might benefit from blocking members of a same category, while participants in the similarity-based order might benefit from interleaving members of different categories. We here address this question, investigating how different types of order interact.

The present study simultaneously manipulated between-category orders (blocked vs. interleaved), within-category orders (rule-based vs. similarity-based), and across-blocks manipulations (constant vs. variable). Manipulations across-blocks were thought to be particularly important because they have not been addressed by previous studies. Our goal was to use the well-studied 5-4 category structure of Medin and Schaffer (1978) to examine the way these different order manipulations interact. This structure has been analyzed in numerous studies (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) and it appeared to be a fruitful baseline 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.

Method

Participants

One hundred and eighty-nine 18-75-year-old participants contributed to this study. One hundred and thirty participants were sophomore or junior students from University Côte d'Azur (France) who received course credits in exchange for their participation. The remaining fifty-nine participants were recruited on Campus on a voluntary basis. Note that the data-set corresponding to the first one hundred and thirty participants has already been used in (Mezzadri, Laloë-Verdelhan, et al., 2021; Mezzadri, Reynaud-Bouret, et al., 2021) for testing categorization models.

Categories

Each participant was administrated a single 5-4 category set (see Figure 1, on the top), composed of $2^4 = 16$ items. In this category set, 5 items belong to category A, 4 items belong to category B, and the remaining 7 items are transfer stimuli.

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. In sum, the task given to participants was unique.

Phases

A learning phase in which participants were instructed to learn the classification of 9 learning items was followed by a transfer phase in which participants were tested upon the presentation of 7 novel stimuli. During learning, two blocks of supervised learning (in which the order of the stimuli was manipulated and feedback was provided) were followed by one block of unsupervised learning (in which stimuli were randomly presented with no feedback). This pattern was repeated until the end of the learning phase. The use of unsupervised random blocks allowed us to assess learning with neither order manipulation nor feedback interfering with the measure of performance. Moreover in a previous work (Mezzadri, Laloë-Verdelhan, et al., 2021), the introduction of unsupervised blocks allowed us to evaluate learning models on the whole learning process instead of exclusively evaluating them on the last learning blocks. The unbalanced ratio of two supervised blocks followed by one unsupervised block 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 unsupervised blocks of 9 stimuli (not necessarily consecutive) to complete the learning phase. Once participants met the learning criterion, the transfer phase was initiated. The transfer phase was composed of 5 blocks of 16 stimuli.

Ordering of stimuli

The experiment was characterized by a full factorial design. Three factors were used, each one having two levels: within-category order manipulation (Rule-based vs. Similarity-based), between-category order manipulation (Blocked vs. Interleaved), and 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, the order was only manipulated in the supervised 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 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. The 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, following these repetitive patterns participants could have guessed the correct classification without paying attention to the stimuli.

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 and B) using these two response keys. Once the key was pressed during the supervised blocks (exclusively), feedback indicating the correctness of participants' classification appeared for 1 s at the bottom of the screen. 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.

Transparency and openness

All data, analysis code, and study materials are available in the Open Science Framework at https://osf.io/w29ts/?view_only=c28b965cc9a74c54b56d7adb87417ff1. Data were analyzed using R, version 3.6.3 ("Holding the Windsock", 2020) and the package ggplot2, version 3.3.3 (Wickham, 2020). This study's design and its analysis were not preregistered.

Results

Learning phase

Because we conducted survival analysis (which accounts for individuals who did not complete the task), none of the participants were removed from the study. An exception to that is the Wilcoxon-Mann-Whitney test that was performed to assess participants' learning time and correct responses. Further details about this test and the number of removed participants will be given in sections "Learning times analysis" and "Correct responses analysis".

Our analysis is threefold: number of unsuccessful participants (i.e., individuals who did not reach the learning criterion), time (in terms of number of blocks) needed by participants to complete the learning phase, and participants' proportion of correct responses. All three measures were assessed across conditions, both taken separately (rule-based vs. similarity-based, blocked vs. interleaved, and constant vs. variable) and combined (R + B + C vs. R + B + V vs. R + I + C vs. R + I + V vs. S + B + C vs. S + B + V vs. S + I + C vs. S + I + V).

Unsuccessful participants analysis

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 maximum amount of blocks allowed to the experiment (200 learning blocks). Table 2 shows the number of successful and unsuccessful participants per type 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-value $= 0.81$ for rule-based vs. similarity-based, p-value $= 0.64$ for blocked vs. interleaved, p-value $= 0.64$ for constant vs. variable, and p-value $= 0.21$ on the table with all conditions). This means that the type of order in which stimuli were encountered did not alter participants' chance of reaching the learning criterion.

Learning times analysis

Three analyses were conducted to determine which condition led to the fastest learning. All three analyses compared the time (in terms of the number of blocks) at which participants reached the learning criterion.

Wilcoxon-Mann-Whitney test. In this analysis, the 20 participants who did not meet the learning criterion were removed. 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). A two-sided Wilcoxon-Mann-Whitney was run to compare the observed learning times. The use of a non-parametric test was preferred to parametric tests such as z- and Student's t-tests to avoid making assumptions about the underlying distribution. The Wilcoxon-Mann-Whitney test was significant at 0.05 level for within-category and across-blocks orders (p-value = 0.027 for rule-based vs. similarity-based, p-value = 0.44 for blocked vs. interleaved, and p-value = 0.02 for constant vs. variable), showing faster learning in both rule-based and constant orders as compared to similarity-based and variable orders, respectively. A statistical analysis of the learning time of participants in each condition will be given in paragraph "Cox proportional-hazards model".

Kaplan–Meier survival curves. The Kaplan-Meier estimator (Kaplan & Meier, 1958) is a survival analysis technique allowing researchers to estimate the expected duration of time until an event of interest occurs (our event of interest is the exact moment of the successful completion of the learning phase). Because this technique takes into account participants who did not meet the learning criterion, all individuals were included in the analysis. Figure 3 shows the survival probability for each type of order taken separately (Figure A) and combined (Figure B) as a function of block number. The survival probability shows how participants assigned to a given condition are likely to continue the task, and consequently 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-value = 0.049, and for constant vs. variable p-value = 0.0076). This shows that learning was faster in the rule-based and constant orders as compared to the similarity-based and variable orders, respectively. The analysis of the combination of the studied types of order is performed in the next paragraph by means of the Cox model.

Cox proportional-hazards model. Similar to the Kaplan-Meier estimator, the Cox model (Cox, 1972) is a survival analysis technique. Again, all participants were included in the analysis. The Cox model is particularly advantageous because of its ability to simultaneously account for multiple variables. Figure 4A shows the result of the Cox model as a function of three variables (within-category, between-category, across-blocks orders). The graph shows that the similarity-based order, the interleaved study, and the variable manipulation across-blocks reduced participants' hazard ratio as compared to their respective reference conditions (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-value = 0.06 for within-category orders, p-value = 0.132 for between-category orders, and p-value = 0.009 for across-blocks orders).

Figure 4B shows the result of the Cox model as a function of the conditions. The hazard ratio of conditions S+B+V and S+I+V were found to be smaller than the hazard ratio of the reference condition R+B+C (p-values = 0.007 and = 0.003, 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 that the combination R+C was the most beneficial, or vice versa, that the combination S+V was the most detrimental.

Correct responses analysis

Because we were interested in analyzing the learning curves of participants who learned the studied categories, unsuccessful participants (amounting to twenty) were removed from the analysis. Figure 5 shows the average percentage of correct responses among participants within the 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.

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 200. As can be seen in Figure 5A, 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 and we corrected the result using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) to take the plurality of tests into account. The test was performed until the block number after which the average percentage of correct responses (in the studied types of order) was higher than 95% (i.e., block number 78 for the within-category orders, block number 75 for both the between-category and across-blocks orders). In the within-category orders, 14 over 26 tests were rejected, whereas in the between-category orders none of the 25 tests were rejected. In the across-blocks order, 12 over 25 tests were rejected. Therefore, the difference between learning curves was significant for both within-category and across-blocks orders. As can be seen in Figure 5B, learning was the less efficient in condition S+I+V followed by condition S+B+V.

Transfer phase

Because we were interested in studying generalization patterns in participants who learned and remembered the studied categories, participants who did not meet the learning criterion (amounting to twenty) as well as participants who incorrectly classified more than 25% of learning items during transfer (amounting to forty) were removed from the analysis. Figure 6 shows the average classification of the transfer items over the course of the transfer phase, as a function of the type of order taken separately (Figure A) and combined (Figure B). Quantity $p(A)$ is the observed proportion that each transfer item was classified into category A during transfer. A two-sided Wilcoxon-Mann-Whitney test was performed to assess the difference in generalization patterns item-by-item. Regarding the within-category order, the test was significant for item T_6 (p-value $= 0.02$). Participants in the rule-based order classified item T_6 into category A less often than participants in the similarity-based order. Regarding the between-category orders, the test was significant for item T_5 (p-value $= 0.046$). However, considering the number of tests that were run (amounting to seven tests per condition), a p-value equal to either 0.02 or 0.046 cannot be considered statistically significant. No difference in generalization patterns was found in participants assigned to across-blocks manipulations.

To further analyze classification patterns, we investigated whether participants in different types of order (taken both separately and combined) applied different strategies to classify new items. Three main strategies were considered: a rule-based strategy that uses Filling pattern (plain vs. striped stimuli) as the main rule, a rule-based strategy that uses Shape (circles vs. squares) as the main rule, and a similarity-based strategy. As mentioned in Section "Ordering of stimuli", Shape as Filling pattern allows participants to minimize the number of exceptions when used as the diagnostic dimension. 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.

Figure 7A shows the putative classification of the transfer items as a function of the applied strategy. For instance, participants adopting a rule-based strategy with Filling pattern as the main rule 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 participants adopting a similarity-based strategy. Therefore, participants who adopted a rule-based strategy with Filling pattern as the main rule could be discriminated from participants who adopted a similarity-based strategy by projecting participants' generalization patterns on Axis 1 = (1,1,0,-1,-1,-1,0). Following similar reasoning, participants who adopted a rule-based strategy with Filling pattern as the main rule could be discriminated from participants who adopted a rule-based strategy with Shape as the main rule by projecting participants' generalization patterns on Axis 2 = (1,1,0,0,-1,-1,0). Likewise, participants who adopted a rule-based strategy with Shape as the main rule could be discriminated from participants who adopted a similarity-based strategy by projecting participants' generalization patterns on Axis 3 = (1,1,0,1,-1,-1,0). Figure 7B shows the nature of the adopted strategy depending on the position of participants' generalization patterns on Axis 1, Axis 2, and Axis 3. For instance, the more participants are on the right side of Axis 1, the more they used a rule-based strategy with Filling pattern as the main rule instead of a similarity-based strategy. Likewise, the more participants are on the right side of Axis 3, the more they used a similarity-based strategy instead of a rule-based strategy with Shape as the main rule.

Figure 8 shows the distribution of participants' generalization patterns on Axis 1, Axis 2, and Axis 3 as a function of the type of order taken separately (Figure A) and combined (Figure B). To facilitate readability, density functions of each type of order taken separately and combined were added. As can be seen in Figure 8A, projections on all three axes of the generalization patterns of participants in the rule-based order were higher than those of participants in the similarity-based order. Similarly, projections on all three axes of the generalization patterns of participants in the interleaved order were higher than those of participants in the blocked order. No striking difference was however found between participants with different across-blocks manipulations. A Wilcoxon-Mann-Whitney test was conducted to assess the significance of the difference between distributions of different types of order (see Table 3). The difference in generalization patterns between participants in the rule-based and similarity-based orders was significant on Axis 1 (p-value = 0.031), meaning that participants in the rule-based order showed more often than participants in the similarity-based order generalization patterns that were more consistent with a rule-based strategy with Filling pattern as the main rule than a similarity-based strategy.

Regarding participants with different between-category orders, the test was significant on all three axes (p-value = 0.034 on Axis 1, p-value = 0.020 on Axis 2, and p-value = 0.032 on Axis 3). This means that participants in the interleaved order showed more often than participants in the blocked order generalization patterns that were more consistent with a rule-based strategy with Filling pattern as the main rule than the studied alternative strategies. Also, participants in the interleaved order adopted more often than participants in the blocked order generalization patterns that were more consistent with a similarity-based strategy than a rule-based strategy with Shape as the main rule.

In Figure 8B, we can notice that the projections on all three axes of the generalization patterns of participants in conditions R+I+C and R+I+V were the highest. The Wilcoxon-Mann-Whitney test was significant in all of the axis (p-value = 0.0032 on Axis 1, p-value = 0.0034 on Axis 2, and p-value = 0.013 on Axis 3, see Table 3). This shows that participants in conditions R+I+C and R+I+V adopted more often than participants in other conditions generalization patterns that were more consistent with a rule-based strategy with Filling pattern as the main rule than alternative strategies. Also, between a rule-based strategy with Shape as the main rule and a similarity-based strategy, participants in conditions R+I+C and R+I+V adopted more often the latter as compared to participants in other conditions.

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 has only focused on a specific type of order manipulation (either within-category or between-category), ignoring potential interactions between different types of presentation order. Here, we manipulated within-category, between-category, and across-blocks manipulation orders within a single task.

Our analysis of the learning phase focused on three aspects: *i)* number of unsuccessful participants, *ii)* block number at which participants met the learning criterion, and *iii)* learning progression. The analysis performed on the number of unsuccessful participants was not significant, thus not showing an impact of the presentation order on the ratio of individuals who did not reach the learning criterion.

The analysis of the learning speed showed that both within-category and across-blocks manipulations influenced the pace at which categories were learned (with the across-blocks manipulation being the main predictor of participants' learning speed). More specifically, the rule-based order was found more beneficial than the similarity-based order, and a constant presentation across blocks was found more beneficial than a variable presentation across blocks. Regarding combined types of order, participants in S+B+V and S+I+V required a higher number of blocks than participants in R+B+C to reach the learning criterion. Therefore, combination R+C was the most beneficial, while combination S+V was the most detrimental.

The analysis of the exact moment at which participants met the learning criterion was complemented by the analysis of participants' learning progression across the learning phase. The analysis of the learning curves showed again that the rule-based order and a constant presentation across-blocks yielded superior learning to that of similarity-based order and to that of variable presentation across-blocks, respectively. Therefore, within-category and across-blocks manipulations affected participants' performance during learning both individually and combined.

Following the interpretation of Mathy and 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' minds. The similarity-based order promotes a structure that is not fully instructive about the categories and thus might temporarily mislead participants, in particular for the 5-4 category set 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 across-blocks presentation 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. Moreover, combining a constant across-block manipulation with a rule-based order might have concurrently promoted the induction of the main rule, thus benefiting learning.

Our analysis of the transfer phase was exclusively focused on participants' classification of the transfer items. We found that within-category and between-category orders had an impact on generalization patterns. More specifically, participants in the rule-based order more often showed generalization patterns consistent with a rule-based strategy with Filling pattern as the main rule than participants in the similarity-based order. Likewise, participants in the interleaved order more often showed generalization patterns consistent with a rule-based strategy with Filling pattern as the main rule than participants in the blocked order. These results show an overall preference for a rule-based strategy with Filling pattern as the main rule in participants assigned to both rule-based and interleaved orders. By contrast, across-blocks manipulations were not found to alter category transfer.

Finally, we found that participants in conditions R+I+C and R+I+V more often adopted a classification logic consistent with a rule-based strategy with Filling pattern as the main rule than participants in other conditions. This shows that the association of the rule-based order with the interleaved order promoted the use of a rule-based strategy based on Filling pattern.

The preference of participants in the rule-based order for a rule-based strategy with Filling pattern as the main rule 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.

The preference of participants in the interleaved order for a rule-based strategy with Filling pattern as the main rule might find an explanation on the Sequential Attention Theory (Carvalho & Goldstone, 2015b). Alternating the presentation of stimuli of different categories might have helped to focus participants’ attention toward the differences between items, promoting the detection of the main rule. Since half of the participants in the interleaved order were concurrently assigned to a rule-based order, it is not surprising that the most favored main rule was the one based on Filling pattern (plain vs. striped stimuli).

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; Meagher et al., 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 researchers to account for individuals who quit the task. We advise the use of similar statistical tools when the conditions allow them.

Study limitations

One limitation of the present study is the use of Medin and Shaffer’s 5-4 category set. These artificial categories are advantageous since they have been thoroughly scrutinized in the past, but they have the disadvantage to be distant from real-world categories. Also, the 5-4 category set has a clear “rule plus exceptions” structure and its ratio between learning and transfer items is particularly high (nine over seven). For these reasons, the present method should be reproduced with different categories and stimuli. Our idea is that we could not use a new structure as a benchmark while adopting both new experimental factors and new statistical tools.

A second limitation is due to a choice regarding the design of the experiment. 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. Indeed, both dimensions minimize the number of exceptions (two in both scenarios) when a “dimension plus exceptions” structure is adopted. Also, 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 the above-mentioned variations would have been too costly.

Acknowledgments

We are thankful to all of the volunteers who took part in 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 article was drafted 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.

Conflict of interest

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

References

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.

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 (2006)*, *65*, 1123–1138.

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.

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

Mezzadri, G., Laloë-Verdelhan, T., Mathy, F., & Reynaud-Bouret, P. (2021). *Hold-out strategy for selecting learning models: Application to categorization subjected to presentation orders* [preprint on webpage at hal.archives-ouvertes.fr/hal-03284595].

Mezzadri, G., Reynaud-Bouret, P., Laloë-Verdelhan, T., & Mathy, F. (2021). *An order-dependent transfer model in categorization* [preprint on webpage at hal.archives-ouvertes.fr/hal-03225670].

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.

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.

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.

Table 1

The 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

Table 2

The 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	9
Similarity-based	85	11
Between-category orders		
Blocked	89	9
Interleaved	80	11
Across-blocks orders		
Constant	88	9
Variable	81	11
Conditions		
R + B + C	21	1
R + B + V	22	3
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

Table 3

Results of the Wilcoxon-Mann-Whitney test on the projections of participants generalization patterns on Axis 1, Axis 2, and Axis 3 as a function of the type of order, taken both separately and combined.

Type of test	Axis	p-value
Within-category orders		
Rule-based > Similarity-based	1	0.031*
Rule-based > Similarity-based	2	0.052
Rule-based > Similarity-based	3	0.150
Between-category orders		
Interleaved > Blocked	1	0.034*
Interleaved > Blocked	2	0.020*
Interleaved > Blocked	3	0.032*
Across-blocks orders		
Constant = Variable	1	0.870
Constant = Variable	2	0.830
Constant = Variable	3	0.500
Conditions		
R + I + C/V > others	1	0.003**
R + I + C/V > others	2	0.003**
R + I + C/V > others	3	0.013*

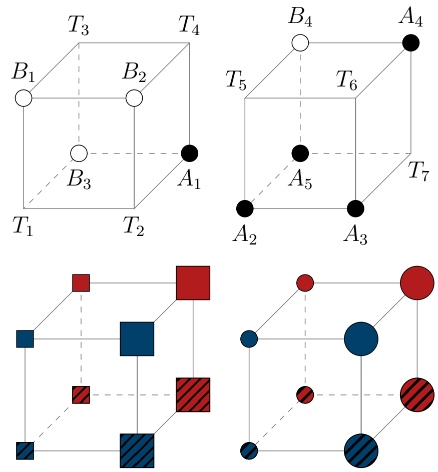

Figure 1

Categories and stimulus items of our categorization task. At the top, the 5-4 category set of Medin and 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).

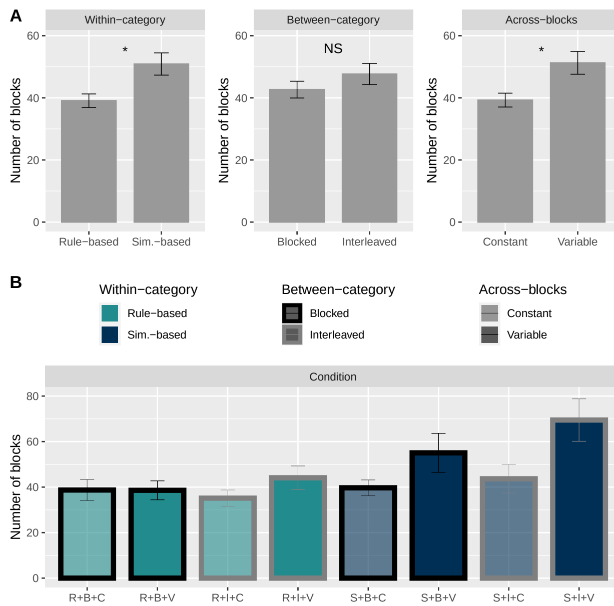

Figure 2

The 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), the 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). Asterisks show the significance of the Wilcoxon-Mann-Whitney test in Figure A. Error bars show $\pm 1SE$.

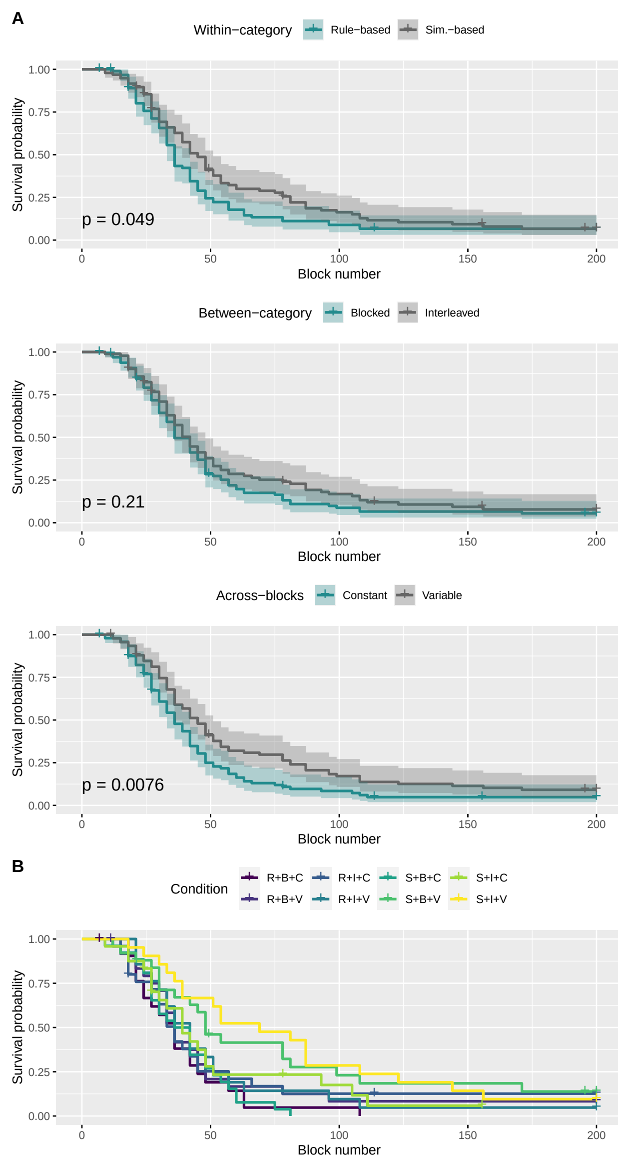

Figure 3

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

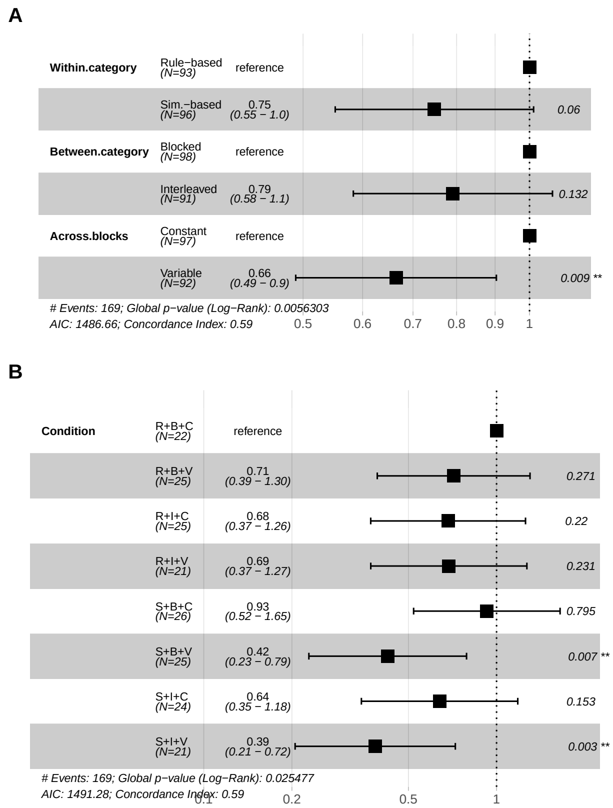

Figure 4

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 for each type of order (taken both separately and combined) in the middle. Statistical significance of the Wald test is shown for each type of order (taken both separately and combined) on the right side.

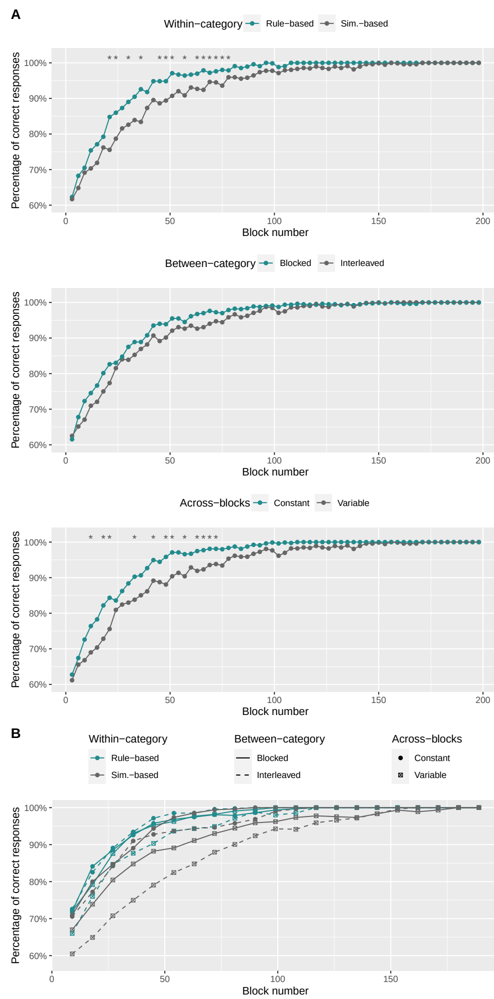

Figure 5

The average percentage of correct responses among participants within the 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 is plotted. Asterisks indicate the blocks on which the Wilcoxon-Mann-Whitney test with the Benjamini-Hochberg correction was rejected. 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 the smoothness of the curves, performance at each random block is obtained by averaging the 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.

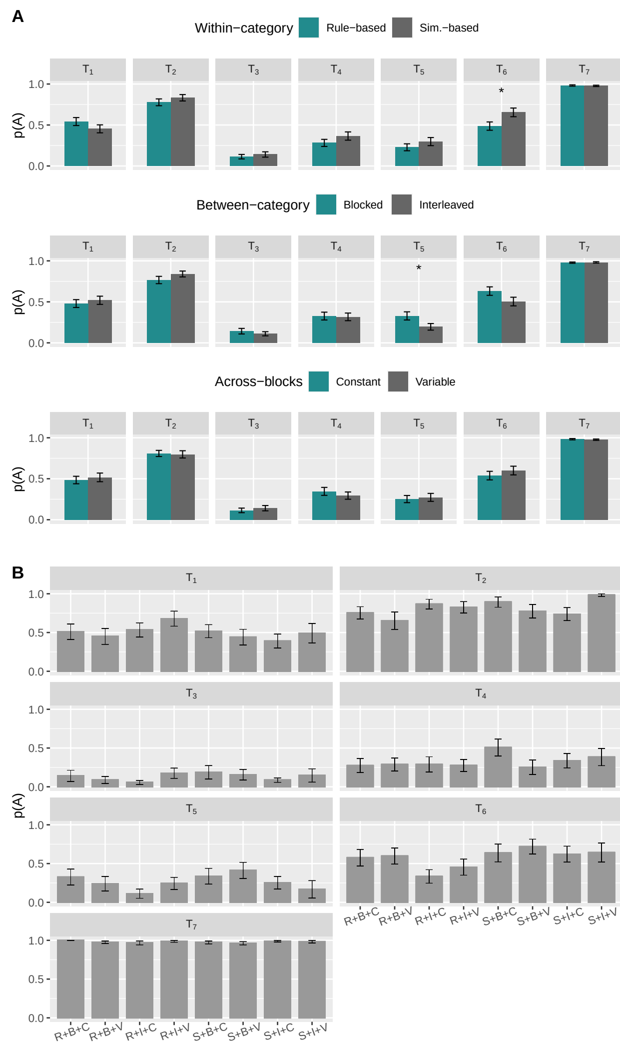

Figure 6

The 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 (Figure A) and combined (Figure B). 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 participant before being averaged across participants. The average classification of the nine learning items is not included. Asterisks indicate the items on which the Wilcoxon-Mann-Whitney test was found significant. Error bars show $\pm 1SE$.

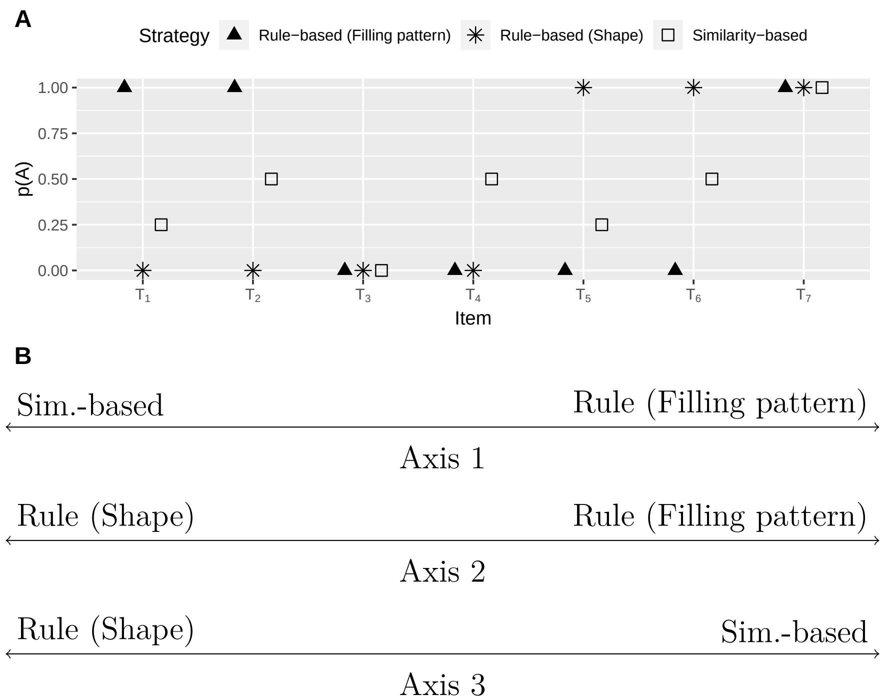

Figure 7

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, and similarity-based 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. On the bottom (Figure B), nature of the adopted strategy depending on the position of participants' generalization patterns on Axis 1, Axis 2, and Axis 3. Coordinates of Axis 1, Axis 2, and Axis 3 are (1,1,0,-1,-1,-1,0), (1,1,0,0,-1,-1,0), and (1,1,0,1,-1,-1,0), respectively.

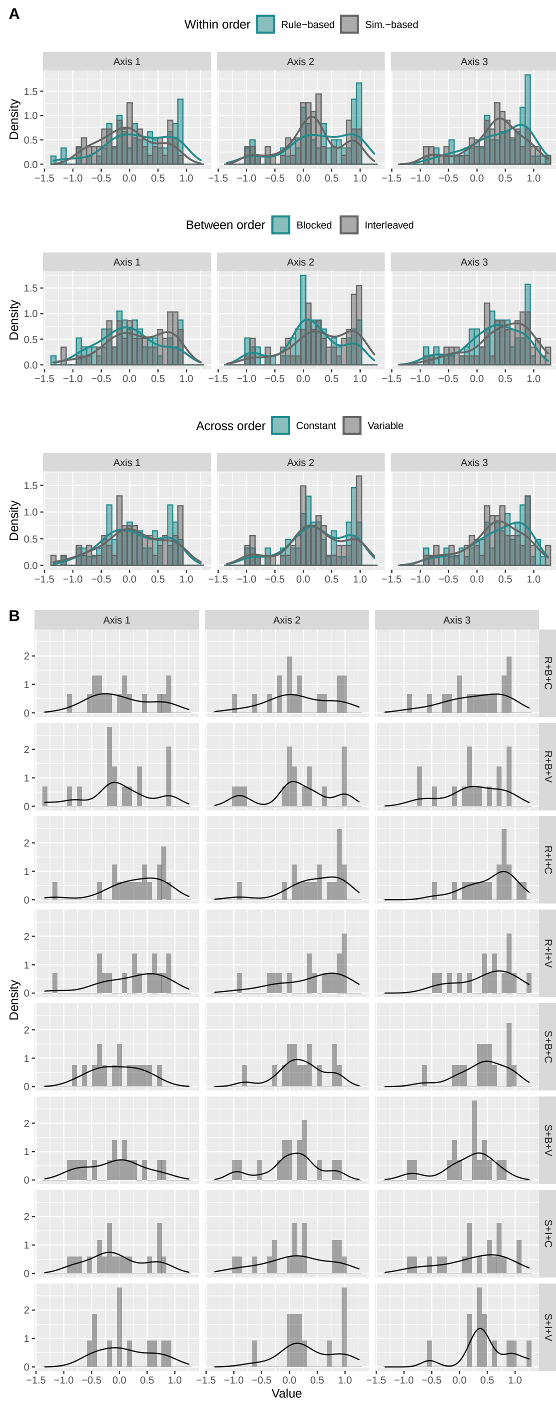

Figure 8

Distributions and density functions of the projections on Axis 1, Axis 2, and Axis 3 of participants' generalization patterns as a function of type of order taken separately (Figure A) and combined (Figure B). Coordinates of Axis 1, Axis 2, and Axis 3 are (1,1,0,-1,-1,-1,0), (1,1,0,0,-1,-1,0), and (1,1,0,1,-1,-1,0), respectively.