Preprint; please don't quote

Serial-order Effects on Category Learning

Fabien Mathy
Université Nice Sophia Antipolis

Jacob Feldman
Rutgers University – New Brunswick

Most category learning models are insensitive to the order in which stimuli are presented, and partly for this reason most category learning experiments have used random presentation orders. However several substantial effects of order have been demonstrated, with some presentation orders leading to markedly faster learning than a random order. In this paper we report several experiments demonstrating and exploring such effects, and introduce a model to account for them. In our experiments, we taught subjects 4-feature category structures using rule-based presentation orders (in which examples drawn from the same regular cluster are presented adjacently in the sequence) and similarity-based presentation orders (which maximize the similarity between successive examples), both of which have been shown to enhance learning relative to random orders. In Exp. 1 negative examples were interleaved within blocks, while in Exp. 2 the chosen presentation order was maintained constant throughout the experiment. Exp. 3 further constrained the constant successive blocks by testing blocked orders in which all the negative examples were segregate from the positive examples. Finally Exp. 4 extended our study of rule-based presentation effects to three canonical category structures from the literature (Shepard Types III, IV, V). To explain the striking effects of serial order found in these experiments, we introduce an extension of the Generalized Context Model (GCM) modified to incorporate temporal information. The model, called Temporal GCM (TGCM), incorporates serial order as a feature along with ordinary features, allowing it to account for the effect of sequential order as a kind of distortion of the feature space. TGCM was able to effectively account for all the effects of presentation order revealed in our data.

Order matters

If you were a realtor, how would you take a customer on a tour of a house? Would you start with the nicest rooms before ending up with the ugliest—based on the idea that the first impressions are primary—or would you choose the opposite tour to leave the potential client with a final good impression? Or would you show the rooms in random order? We believe that these alternative sequences inevitably leave a different mental representation of the house. Although timing and presentation orders have been known to effect retention (as a function of the manipulation of study events; see Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008a), diagnosis on the basis of a simple manipulated checklist of symptoms (Kwan, Wojcik, Miron-Shatz, Votruba, & Olivola, 2012), or misidentification in lineups (Wells, 2014), studies on categorization usually have made every effort to randomize trial order. Thus with the exception of studies surrounding the few categorization models that are known to implement an incremental architecture (Love, Medin, & Gureckis, 2004; Sakamoto, Jones, & Love, 2008; Stewart, Brown, & Chater, 2002) which have touched on presentation orders, the vast majority of studies draw the to-be-categorized stimuli in random order in order to avoid a list of potentially confounding variables. As a result, the influence of presentation order is understudied and scarcely addressed at all in standard models (Markant & Gureckis, 2014; McDaniel, Cahill, Robbins, & Wiener, 2014; Rouder & Ratcliff, 2004).

But order does matter in categorization, as it does in other domains such as short-term memory (Farrell, 2008, 2012; Miller & Roodenrys, 2012). Recent evidence suggests that the representations formed during a categorization process can be distorted by an interleaved vs. blocked sequential order of the presented stimuli (Birnbaum, Kornell, Bjork, & Bjork, 2013a, 2013b; Clapper & Bower, 1994; Carvalho & Goldstone, 2014; Goldstone, 1996; Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010; Kang & Pashler, 2012; Wahlheim, Dunlosky, & Jacoby, 2011). In addition to a spacing effect (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008b; Hintzman, Summers, & Block, 1975), interleaving the stimuli of different categories has been shown to highlight their differences (Birnbaum et al., 2013a), although the manipulation of similarity can modulate this advantage (Carvalho & Goldstone, 2012; Weitnauer, Carvalho, Goldstone, & Ritter, 2013). The interleaved vs. blocked factor has extended previous research showing within-category order effects on category learning (Elio & Anderson, 1981, 1984; Medin & Bettger, 1994; Stewart et al., 2002; Skorstad, Gentner, & Medin, 1988).

This research was supported in parts by the Agence Nationale pour la Recherche (ANR) Grant # ANR-09-JCJC-0131-01 to Fabien Mathy and by NSF EITM 0339062 and NIH EY021494 to Jacob Feldman. We are grateful to Azizedine Elmahdi and Nicolas Heller for assistance in data analysis. Correspondence concerning this article should be addressed to Fabien Mathy, BCL Lab, CNRS, UMR 7320, Université Nice Sophia Antipolis, France. or by e-mail at fabien.mathy@unice.fr.

Medin & Bettger, for example, have shown that maximizing the similarity between successive examples within categories leads to more efficient learning. The present study describes both between-category (interleaving vs. blocking) and within-category presentation order effects (using different types of orders, including similarity), and also investigates constant vs. variable presentations across blocks. Although these previous studies have shown a variety of order effects, almost none of them has attempted to model precisely how the contiguity of the stimuli can affect one another generally. One exception is the distributed temporal context model (TCM) by Howard and Kahana (2002) intended to model a different but related problem, recency and contiguity effects in serial and free recall. We here present a different context model which is more minimalist and which can be adapted in a more straightforward fashion to the categorization domain in order to account for recency, contiguity, primary effects, and also spacing.

Rule- vs. Similarity-based presentation orders

One goal of manipulating presentation orders is to help understand whether learners use rules, exemplars or both to mentally represent categories (Allen & Brooks, 1991; Ashby & Ell, 2001; Goldstone, 1994; Hahn & Chater, 1998; Homa, Rhoads, & Chambliss, 1979; Komatsu, 1992; Pothos, 2005; Rips, 1989; Rosch & Mervis, 1975; Rouder & Ratcliff, 2004; E. E. Smith, Patalano, & Jonides, 1998; E. E. Smith & Sloman, 1994; Thibaut, Dupont, & Anselme, 2002; Thibaut & Gelaes, 2006). Based on these two dominant theoretical views, Mathy and Feldman (2009) devised different paradigm-based within-category presentation orders and found that a *rule-based* presentation order leads to faster learning in comparison to a *similarity-based* order, which prior to this result had been the only order known to facilitate learning in artificial classification tasks (Elio & Anderson, 1981, 1984; Medin & Bettger, 1994). In the rule-based order, the stimuli were ordered following a rule-plus-exception structure, meaning that examples obeying a common rule were presented adjacently in the sequence (in random order), separated from "exceptions" (examples that don't follow the principal rule). This type of order seems to aid learning because it facilitates the induction of the relevant rules. The rule-based order generally results in low similarity between adjacent examples, in part because examples within a "rule" are presented in random order. In contrast, a similarity-based order is designed to maximize the similarity between contiguous examples. Based on the results, we concluded that the rule-based order offers a learning advantage beyond that provided by inter-item similarity per se.

However, the mechanisms underlying these order effects are still unclear. On the one hand, one might imagine multiple mechanisms, with rule-based orders facilitating rule-learning processes, and similarity-based orders facilitating similarity-based mechanisms such as those at work in exemplar models. On such an account, two distinct mechanisms of learning would coexist (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), each benefiting from presentation orders consistent with their respective biases. Along these lines, our own later studies (Mathy & Feldman, submitted) suggest that a rule-based training condition leads to generalization patterns consistent with rule-based retrieval, while a similarity-based training condition shows generalization patterns more consistent with exemplar retrieval.

On the other hand, one would like to arrive at a unified account that explains performance in all conditions, and indeed it is not hard to imagine that a single account can explain the observed differences. For example, a rule-construction mechanism might explain both the overall superiority of rule-based orders as well as the advantage for similarity orders relative to random orders. On such an account, the benefit of the similarity order would reflect the tendency for similar items to obey common rules, while the even greater benefit of the rule-based order would reflect the alignment between the order of the examples and the types of rules towards which the mechanism is biased. On this latter account, sequences of similar examples might occasionally "mislead" the rule-extract mechanism, suggesting regularities that are actually artifacts of the similarity order. An example would be a sequence of red examples that suggests a "red things" category, when in fact no such rule exists in the target concept. A rule-based order, by its construction (random order among items that obey a common rule) tends to defeat such accidental alignments.

Nevertheless, even if there is a single common mechanism, it is by no means clear that it is "rule-based." In the current paper, we pursue a particularly parsimonious hypothesis in which presentation order effects are explained by a common similarity-based (exemplar-based) mechanism, but one in which—unlike conventional exemplar models—sequential order is incorporated as a feature.

Our approach

The present investigation began with the premise that an exemplar model, if suitably modified, might account for presentation order effects. As we will detail below, such models provide a simple avenue for incorporating temporal order without any major modification in their architecture—namely, by including temporal "distances" in a like manner to the feature distances ordinarily used in similarity-based models. These temporal distances then distort the stimulus space via the conventional similarity-based mechanisms of the General Context Model (Nosofsky, 1984, 1986), producing effects of stimulus order. To test this model, we carried out a series of experiments investigating presentation order effects in more detail than has been done in previous studies, including several novel variations in the way the presentation orders were constructed. To preview, our participants were given rule-based category structures that could be learned via some explicit reasoning process, and therefore we naturally expected to show a positive effect of rule-based presentation orders on learning as already shown in Mathy and Feldman (2009). However, instead of simply confirming the obvious (e.g., a rule-based model predicts better learning when presentation order is rule-based), our aim is to incorporate all our results into a single unified model.

Overview of experiments and words of caution

In Mathy and Feldman (2009), we used presentation orders that were variable across blocks: first, each new block was newly randomized (within the constraints of the respective desired order) in order to avoid the repetition of identical sequences of category responses across blocks. Second, we randomly intermixed the negative examples among the positive ones, which somewhat diluted the desired order. Third, the negative examples were not ordered in spite of the possible effect of their presentation on rule-abstraction or exemplar memorization. These choices can presumably complicate the modeling of presentation-order effects. In the present study, the participants in the first experiment were each given a series of variable orders across blocks to replicate Mathy and Feldman (2009) while using a procedure that was more similar to the two other experiments of the present study. In the second experiment, the orders were maintained constant within participants, but they varied between participants, with the negative examples randomly intermixed between participants (e.g., + + - + - + + + - - + - - - + -). Because this procedure could presumably help the participants perceive sequential subpatterns of responses (e.g., + - + -) that could be used as a way to avoid the targeted visual classification, this experiment was only briefly tested with a small sample of participants. To fix orders in a maximal way, a more important third experiment used fully-blocked orders across blocks (the negative examples were all presented after a separate block of positive examples, + + + + + + + + - - - - - - - -, and order within-category was set constant). This last condition required separating the training blocks (in which the presentation order was manipulated, but in which successive identical classifications would guarantee perfect performance if the participants were required to make a classification) from the categorization blocks (in which the presentation order was randomized to have the participants make a classification). The training blocks and the categorization blocks were run alternatively, and the participants were expected to learn the categories during the training blocks and to perform during the categorization blocks.

We hypothesized that constant orders between the successive learning blocks would enhance rule-based presentation order effects by allowing the participants' hypotheses to be tested in a more stable context (for example, seeing a positive large red square followed by a negative small red square several times might help induce a rule on the Size dimension). A second hypothesis was that rule-based blocked presentations would favour the abstraction of commonalities within categories (by opposition to intertwining the negative and positive categories), which was hypothesized to benefit rule-based learning. In both cases (constant orders and blocked presentations), it was hypothesized that exemplar-based representations would also benefit from the reinforced associations caused by the reduction of the number of temporal associations using constant orders, and that blocked orders would reinforce associations within categories. The manipulation of orders (variable between blocks, constant between blocks, or fully blocked) were therefore expected to have the same progressive positive effets on rule-based and similarity-based orders.

In three experiments, a single rule-based categorization task identical to the most complex category structure studied in Mathy and Feldman (2009) was used to allow some comparisons between experiments. The fourth experiment used a set of more standard classification tasks, Types III-IV-V of Shepard, Hovland, and Jenkins (1961), to show how rule-based blocked orders can be beneficial to learn some easier category structures and to show that quick effects can be expected from manipulated orders on smaller samples of stimuli. Note that although different category structures are known to modulate the advantage of certain orders (Carvalho & Goldstone, 2014), in the first three experiments, we chose to vary the procedures instead of the category structures, and to study the cumulative effect of constant and blocked presentations. This choice did not allow us to study the additive and interactive effects of the present factors at play. Therefore, our findings are only informative about the utility of manipulating orders to test a categorization model, but other effects could possibly be observed with other category structures for the same presentation orders, and simulation could react in a different fashion as well.

One final word of caution is required. Our experiments do not attempt a comprehensive comparison of all possible presentation orders, but rather focus on a particular set of theoretically revealing comparisons. A full comparison would in any case be impractical, in that it would require $2 \times 2 \times 4 = 16$ experimental conditions to study the separate and combined effects of constant vs. variable orders, blocked vs. interleaved orders, and type of orders (rule-based, similarity-based, dissimilarity-based, and random), and 24 conditions to further differentiate interleaving vs. alternating the negative examples, not to mention the three- versus four-dimensional settings that would double the 16 or 24 conditions. We generally focus on contrasting rule-based and similarity-based orders (under a variety of conditions), which have both been found to improve learning relative to dissimilarity-based and random orders. In Exp. 4

we use a smaller stimulus set than Exp.1-3. As this could potentially diminish the superiority of the rule-based order over similarity-based orders, we decided to contrast the rule-based-order with the dissimilarity-based order (which has been shown to be particularly detrimental to learning) to increase the chances of observing a significant difference between the order types. Once again, our overall goal is to demonstrate a number of empirical effects of presentation order and bring them all together under a unified theoretical account.

Experiment 1

This experiment is closely based on that of Mathy and Feldman (2009), modified in several ways to better match the methods of Exps. 2 and 3. In particular, while in Mathy and Feldman (2009) we manipulated only the order of the positives, here we control the order of both positives and negatives; and here only one concept was given to the participants instead of two as in the earlier study. In addition, whereas Mathy and Feldman (2009) only analyzed data from participants who could solve the classification problem up to a fixed learning criterion, here we included all participants' data in the analysis in order to get a more comprehensive view of performance. As a consequence the mean level of performance is naturally lower.

Method

Participants. The participants were 68 freshmen and sophomores at the Université de Franche-Comté (France) who received course credit in exchange for their participation.

Procedure. There was no warmup session (such as learning a simple one-dimensional concept) so that the participants would not think that the tasks consisted in searching for simplistic rules. However, the participants were briefed before the task began. Each participant was asked to learn a single concept (detailed below) and was given different presentation orders across blocks, but each participant was given one type of order (rule-based, $N = 34$, or similarity-based, $N = 34$).

The task was computer-driven and the participants were tested individually during a one-hour single session (including briefing and debriefing). The participants sat approximately 60 cm from a computer on which stimulus objects were presented one at a time in the upper part of the screen. They learned to sort the stimulus objects using two keys, and successful learning was encouraged by means of a progress bar. The positive and negative categories were associated with the up and down keys respectively, and by two category pictures on the right hand side of the screen. A virtual frame for the categories faced the frame that encompassed the stimulus on its left. The frame for the categories displayed a schoolbag at the top, and a trash can at the bottom (to match the response keys). Each time a response key was pressed, the corresponding picture was displayed for two seconds along with feedback, while the opposite picture was hidden for two seconds. After each response, feedback indicating a correct or incorrect classification was given at the bottom of the screen for two seconds. The two category pictures reappeared whenever a new stimulus was presented.

The participants scored one point for each correct response which was shown on the progress bar. To regulate the learning process, each response had to be given in less than eight seconds (resulting in a maximum of 10 seconds between two stimuli when the participants got a 'Too late' message that lasted two seconds). If the response was given too late, the participants would lose three points on the progress bar. This was thought to prevent the participants from skipping the most difficult stimuli without any penalty. The number of empty boxes in the progress bar alloted to learning was 4×2^D (D = number of dimensions, which was equal to four in our study). One empty box was filled whenever a correct response was given, but the progress bar was reset in case of an incorrect response. This criterion was identical to the one used by Shepard et al. (1961) in their first experiment and by Mathy and Feldman (2009). Consequently, the participants had to correctly classify stimuli on four consecutive blocks of 2^D stimuli to be allowed to stop the experiment. This required the participants to correctly classify all the stimuli, including those considered as exceptions (in accordance with a rules terminology that we use below), and intentionally limited the participants from adopting strategies such as providing partial solutions (akin to classifying stimuli on the basis of a limited number of features with less than 100% accuracy).

Choice of concepts studied. Each participant was given a concept defined over four Boolean dimensions. We chose to restrict our experiment to the most complex concepts studied by Mathy and Feldman (2009). According to the classification of Feldman (2003), this concept is called $12_{4[8]}$ (Fig. 1) to indicate that it is the 12^{th} in a set of 4-dimensional concepts consisting of 8 positive examples. As described below, this concept presents an interesting set of clusters, but mainly, the choice of a four-dimensional concept can easily be justified by the fact that the number of objects to be classified ($2^4 = 16$) is large enough to bring out any effect of presentation order, but small enough to be learnable.

This concept can be defined by the compressed formula $12_{4[8]} \cong a'(bc)' + ad'(bc' + b'c)$, a useful representation of the clusters that compose $12_{4[8]}$. We use a standard notation here (Feldman, 2000, 2003) in which a' refers to negation ($\neg$) of feature a (a and a' are the two dimension values that can be taken by dimension A), ab refers to the conjunction ($\wedge$)

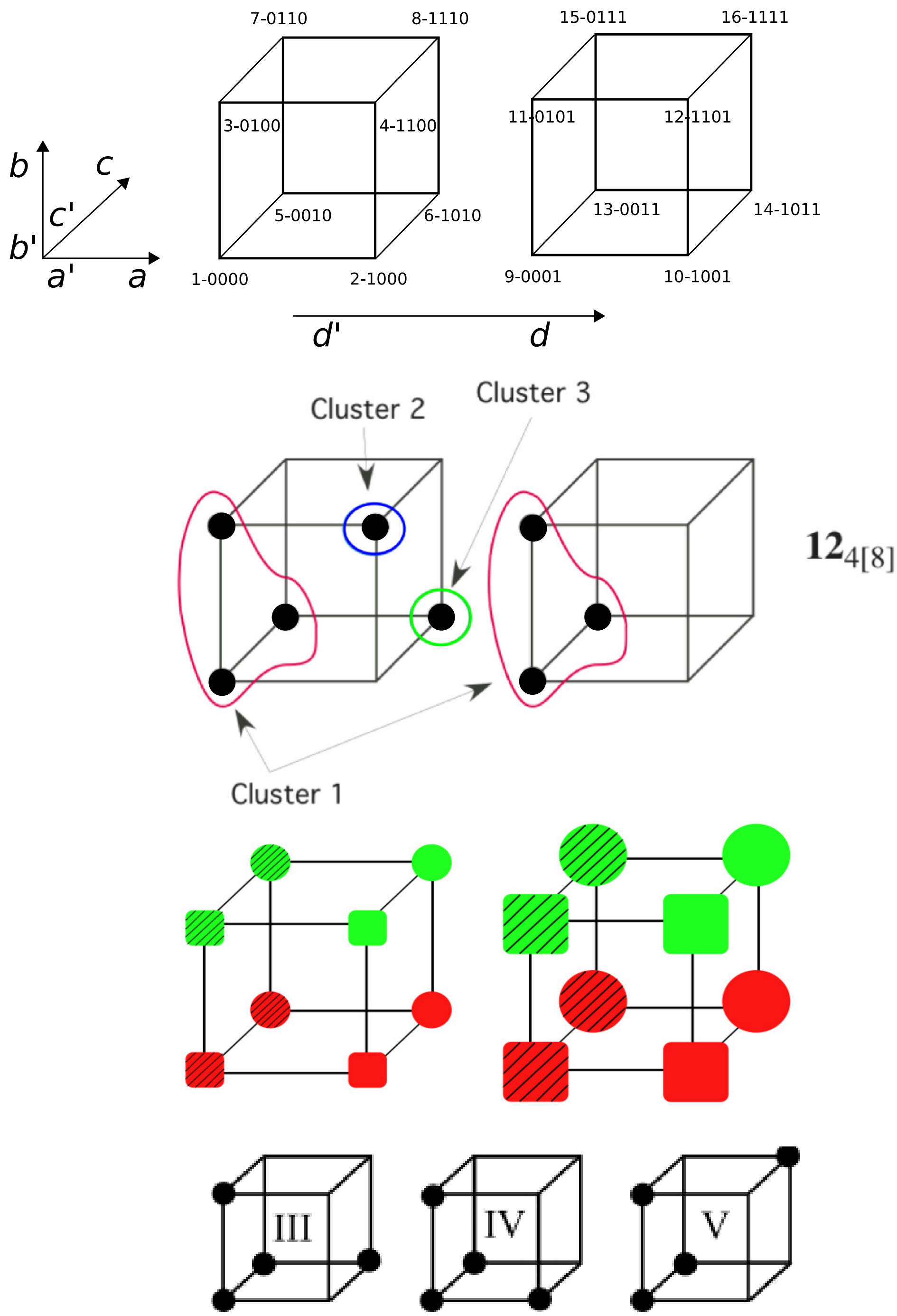

Figure 1. Concept $12_{4[8]}$ and Types III, IV, and V. *Note.* In the $12_{4[8]}$ notation, the [8] extension means that there are 8 positive examples in the concept, 4 means that the concept is four-dimensional, and 12 is an arbitrary label which identifies this concept from among the other concepts available within the 4[8] set (Feldman, 2003). The positive examples of a given concept are indicated by black circle markers in the hypercube, whereas negative examples are represented by empty vertices. The examples are also all listed in Table 1. Among the positive examples, there are three clusters in concept $12_{4[8]}$. In the hypercube representing $12_{4[8]}$, the red set indicates the positive objects belonging to the biggest cluster, the blue set indicates the positive objects of the second cluster, and the green to those from the third one. The clusters of the negative categories are not represented to lighten the presentation. The stimulus coding order is *ABCD*, each of the uppercase letter representing one dimension. The code 0000 stands for $a'b'c'd'$, 1111 stands for *abcd*, each of the lower case letters representing a dimension value (i.e., features). The number preceding the code (1,...16) is a simpler identification number. This concept can be defined by the compressed formula $12_{4[8]} \cong a'(bc)' + ad'(bc' + b'c)$. In the III, IV, and V Types, only the large shapes are considered to form a set of eight stimuli. The following rule-exception separation was defined as follows for the particular cubes in this figure: Hatched except(green circle) or plain red circle for Type III; Hatched except(green circle) or plain red square for Type IV; Hatched except(green circle) or plain green circle for Type V). We also considered an alternative simpler rule for Type III (Hatched square or red circle), by ordering the examples in the rule-based condition in a way that would make the two kinds of abstraction possible: for instance, hatched red square, hatched green square, hatched red circle, and plain red circle can satisfy both kinds of abstraction.

of a and b, and $a + b$ to their disjunction ($\vee$). The $\cong$ symbol indicates that any other concept isomorphic to this formula can be labelled $12_{4[8]}$. For instance, $1_{4[8]} \cong a$ indicates the set of concepts made of structures equivalent to a by rotation or mirror reflection, that is a', b, b', c, c', d, or d' (in each case, exactly one feature defines the concept). The formula for $12_{4[8]}$ is compressed in that it faithfully represents the set of eight positive members of the concept (which unpacked give: $a'b'c'd'$, $a'bc'd'$, $a'b'cd'$, $abc'd'$, $ab'cd'$, $a'b'c'd$, $a'bc'd$, and $ab'c'd$). Following the Port-Royal terminology, the formula can be thought as a definition of the concept, and the unpacked examples as the extension of the concept.

The top of Fig. 1 shows a four-dimensional hypercube made of $2^4 = 16$ stimuli encoded from 0000 (standing for $a'b'c'd'$) to 1111 (standing for $abcd$). The $12_{4[8]}$ concept is shown in an arbitrary rotation in the second hypercube of Fig. 1, in which the eight positive examples are indicated by black circle markers. These hypercubes, also called Hasse diagrams, are extremely useful as a way of visualizing conceptual structures, which do not easily pop up in the corresponding truth tables.

The reasons why this concept has interesting properties are detailed in Mathy and Feldman (2009). In summary (1) the concept is moderately complex, and (2) the concept has a substructure made of several well-defined clusters. Cluster 1 represents six of the concept's eight members, corresponding to the first disjunctive clause $a'(bc)'$ in the concept's compressed formula. These six objects can collectively be represented by a verbal expression such as "all a' except bc". These objects are delimited in red in Fig. 1. By contrast, Clusters 2 and 3 consist of only one object each, each one requiring four literals in order to be identified (respectively $abc'd'$ and $ab'cd'$, corresponding to the expansion of the second clause in the formula, and requiring each four features to be represented verbally). Thus Cluster 1 plays the role of a salient "rule", while Clusters 2 and 3 play the role of "exceptions". Following Mathy and Feldman (2009), we hypothesize that splitting up $12_{4[8]}$ into these particular clusters is beneficial to learning.

Stimuli. Stimulus objects varied along four Boolean dimensions (Shape, Color, Size, and Filling texture). Rotation and permutation were randomized in our experiment for each participant, meaning that dimension A could be one of the Shape or Color dimension, etc., and that features within dimensions were randomly drawn and permuted (for instance, $a' = blue$ and $a = red$, or $a' = red$ and $a = blue$, or $a' = green$ and $a = red$, etc.). The choice of two values for each feature was chosen at random (shape = triangle, square, or circle; color = blue, pink, red, or green; filling = hatched or grilled; size = small or big). Overall, the combination of these four separable dimensions (Garner, 1974) formed 16 single unified objects (e.g., a small hatched red square, a big grilled blue circle, etc.). The use of such monolithic stimuli is particularly helpful to avoid numerical biases that may arise when as many dimensions as objects are used to create stimuli (Mathy, 2010).

Ordering of stimuli. The two presentation orders that best facilitated learning a rule-based order and a similarity-based order were retained from the study by Mathy & Feldman. Presentation order was a between-subject manipulation. One presentation order type was randomly chosen for a given participant beforehand and then applied across the blocks. However, a new presentation order was computed before each block. As in Mathy & Feldman, each new block (although constrained to a given order type) was newly randomized. The negative examples were also randomly intermingled with the positive ones in order to avoid long uninterrupted sequences of positives and negatives. The random organization of the negative and positive examples was also randomized across blocks.

Unlike the study by Mathy & Feldman, the negative examples were also clustered. The second cluster was defined by the negation of $(bc)'$ on the a' dimension, that is, $a'((bc)')'$ or simply $a'bc$, comprising the examples 0110 and 0111. The first cluster was defined by the negation of $d'(bc'+b'c)$ on the a dimension, that is, $a(d'(bc'+b'c))'$, comprising the rest of the negative examples. Therefore, the negative clusters were simply regarded as an inversion of the positive clusters.

In the **rule-based order**, the positive objects were randomly drawn from Cluster 1 until all 6 of them were presented. Likewise for the negative objects belonging to Cluster 1. These were followed by the positive objects in Cluster 2 and Cluster 3 (in random order), and by the negative objects belonging to Cluster 2 (in random order). Thus in the rule-based order, all the members of the biggest cluster per category were presented first, in random order, and separated from exceptional members, in order to promote the abstraction of the simplest rules by participants. The presentation within clusters was randomized to obey a rule-abstraction process that is supposed to impede stimulus singularity.

In the **similarity-based order**, the first object was chosen at random, and subsequent objects were chosen randomly from those maximally similar to the previous object, and so forth until the set of examples was exhausted. The negative examples were also similarity-based ordered. Ties were resolved randomly. Similarity was computed on a trial-by-trial basis so as to maximize inter-item similarity locally, a method which did not guarantee a maximized inter-item similarity over an entire block, but which offered a greater number of possible orders. Dissimilarity between two stimuli i and j was computed with the Minkowski metric

$$d_{ij} = [\sum_{a=1}^{n} |x_{ia} - x_{ja}|^r]^{1/r} \quad (1)$$

where x_{ia} is the value of stimulus i along dimension a. We used a city-block metric appropriate to the separable dimensions used in this study ($r = 1$). The similarity was simply computed using $s_{ij} = n - d_{ij}$. The most important aspect of this procedure is that the ordering does not necessarily respect the cluster boundaries targeted in the rule-based order, as similarity steps can cross in and out of clusters. For instance, the stimulus 0100 can be followed by stimulus 1100.

These two presentation orders match two extreme ways of learning: a complex inductive process based on abstraction and an elementary process with underlying associative mechanisms. The latter mechanism will be targeted in the subsequent simulations.

1. The rule-based condition makes use of a set of clusters which are presented to participants in an order depending on their magnitude (since in many domains, exceptions are learned last), with no distinction within clusters (since the abstraction process is supposed to annihilate any effect of non diagnostic features on learning). Because the objects are supposed to entail common abstract properties within clusters, they are randomly drawn. This presentation order is supposed to facilitate learning given the hypothesis that participants naturally follow an identical rule-based strategy.

2. The similarity-based condition tends to follow an exemplar model, in which there is no specification of how presentation order could affect performance. Effectively, proponents of exemplar models assume that classification is only determined by the degree of similarity between a stimulus and the stored exemplars. A direct consequence of this view is that the computation of similarity in some exemplar models is batched, that is, not trial-by-trial but for the whole set of exemplars (although connectionist exemplar models such as ALCOVE and AMBRY that involve trial-by-trial updating of their connections might better modulate the computation of similarity; see Kruschke, 1992, 1996). Because the exemplar model proceeds by simple associative mechanisms, the temporal contiguity of the stimuli might reinforce (hypothetically) memory traces locally and result in faster learning.

Table 1 shows one similarity-based-ordered block in the SBO-Exp1 column, sampled from the set of presentation orders that could be generated by our procedure. The RBO-Exp2 columns exemplifies one rule-based presentation order for Exp. 1, except that the orders given to one participant were subject to variability across blocks (which was not the case in Exp. 2 in which the computed order was set to remain constant across blocks for one participant).

Results

The mean inter-item similarity for each condition and each experiment is given in Table 2. As expected, the average inter-item similarity was higher for the similarity-based orders than for rule-based orders. The learning curves in Figure 2 *A* show the influence of presentation order on learning and represent the proportion correct across the first 40 blocks. Confirming the result of Mathy and Feldman (2009), learning appears faster in the rule-based order when the number of blocks exceeded 25, but a visual inspection of the plots also confirms that learning was slower in the present experiment than in Mathy and Feldman (2009). One obvious reason is that the participants in the 2009 study were selected for the statistical analysis only if they had reached the 100% criterion. Instead of obtaining more than 85% correct responses around 20 blocks as did the participants in 2009, performance is still around 75% in the present experiment, with the rule-based and similarity-based conditions still undifferentiated at this point. A clearer demarcation between the curves appears around 30 blocks.

We conducted a mixed-model analysis to study the effect of the presentation order on the mean proportion correct, which is shown in Figure 2 *A*, with Block number as the repeated measure and Presentation order as a fixed factor. The mixed-model analysis was thought to facilitate further tests to be carried out in the Results section of Exp. 2 and Exp. 3, in view of studying the variable vs. constant vs. blocked presentation effects and their interaction with order types. In the present experiment, the mixed-model was only used to test the fixed effects of Presentation order, Block number and their interaction. The results showed that Presentation order ($F(1, 2347.9) = 16$, $p < .001$) and Block number ($F(39, 138.8) = 38.9$, $p < .001$) each had a signicant effect on performance. A significant interaction ($F(39, 138.8) = 1.6$, $p = .03$) between the two factors indicated that the curves separated progressively as shown in Figure 2 *A*. The estimates of the fixed effects of presentation orders showed a 95% condence interval that the improvement seen in the rule-based condition was between .02 and .20 more than for the similarity-based condition (the mean effect was equal to .11 across blocks).

Experiment 2

Experiment 2 is designed to investigate the effect of constant orders, meaning orders that are consistently repeated over blocks. We hypothesized that such an order could facilitate both the perception of commonalities within categories (when two stimuli of the same category are presented repeatedly and contiguously) and the perception of contrasts between categories (when two stimuli of different categories are presented repeatedly and contiguously) in order to form an abstraction. From an exemplar point of view, constant orders were thought to limit the number of temporal associations between exemplars, which should therefore reinforce the limited set of associations between the memory traces.

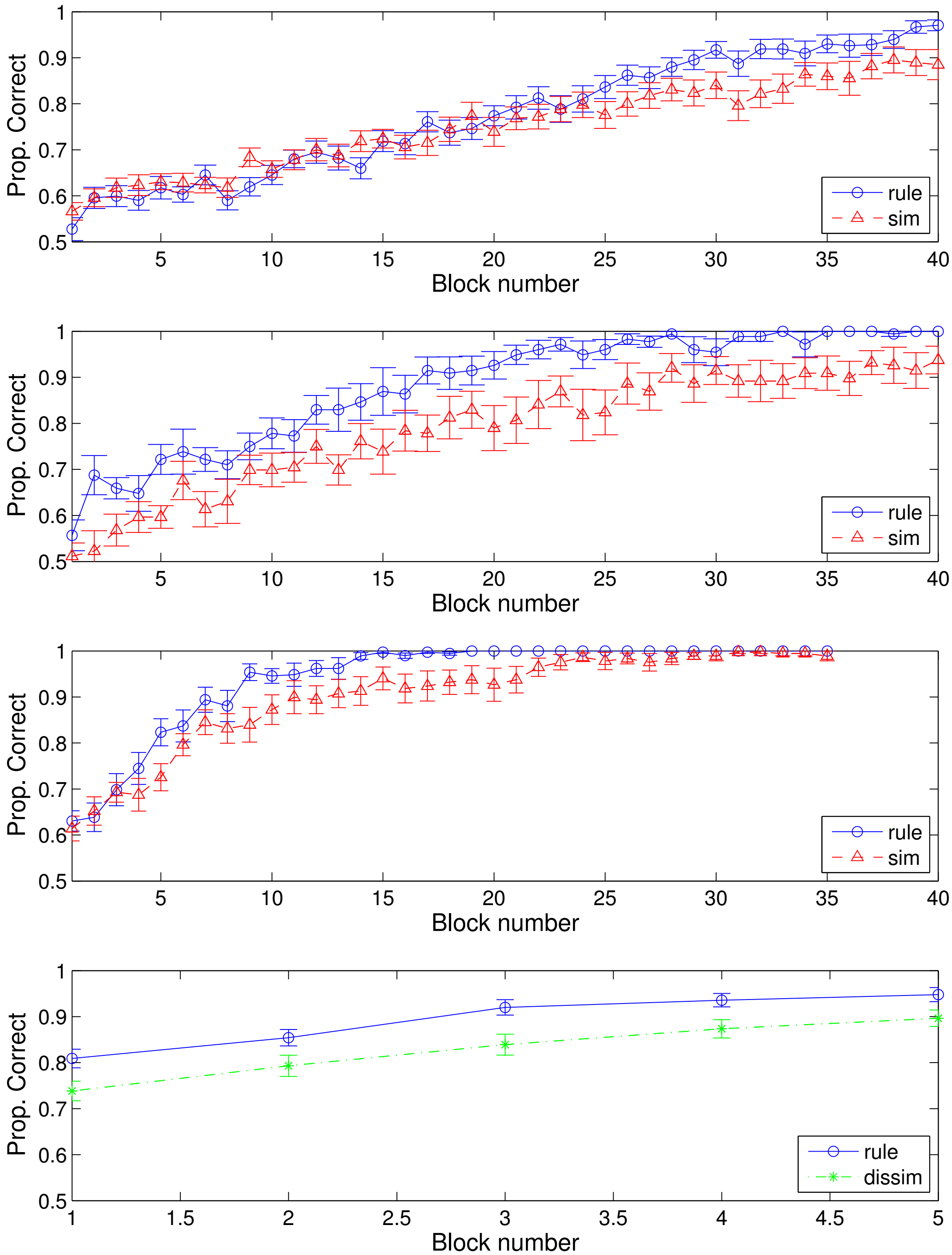

Figure 2. Proportion correct as a function of block number. *Note*. A) Exp. 1, using concept $12_{4[8]}$, with categories interleaved and variable orders across blocks B) Exp. 2, using concept $12_{4[8]}$, with constant orders across blocks and interleaved categories C) Exp. 3, using concept $12_{4[8]}$, with blocked constant orders across blocks D) Exp. 4, using Types III, IV and V, with categories intermixed and variable orders across blocks. Blue curve and circle markers, rule-based order; Red curve and triangle markers, similarity-based order; Green curve and star markers, dissimilarity-based order.

Table 1
Encoded study items of Concept $12_{4[8]}$ *presented in Fig. 1, and three presentation order sampled from the three first experiments.*

$12_{4[8]}$			Presentation order samples in $12_{4[8]}$					
#	Cat A		SBO-Exp1		RBO-Exp2		RBO-Exp3	
1	0000		3	0100	3	0100	3	0100
3	0100		4	1100	13	0011	13	0011
4	1100		7	0110	12	1101	1	0000
5	0010		6	1010	1	0000	5	0010
6	1010		8	1110	14	1011	11	0101
9	0001		5	0010	5	0010	9	0001
11	0101		1	0000	11	0101	4	1100
13	0011		9	0001	9	0001	6	1010
			16	1111	8	1110	12	1101
#	Cat B		14	1011	4	1100	14	1011
2	1000		11	0101	2	1000	8	1110
7	0110		13	0011	16	1111	2	1000
8	1110		10	1001	10	1001	16	1111
10	1001		2	1000	6	1010	10	1001
12	1101		12	1101	7	0110	7	0110
14	1011		15	0111	15	0111	15	0111
15	0111							
16	1111							

Note. Cat A, positive objects of the concept; Cat B, negative objects. The $12_{4[8]}$ concept is shown in Fig. 1; SBO-Exp1, Similarity-Based Order sampled from the many different orders that could be instanciated in Exp. 1; RBO-Exp2, Rule-Based Order for one participant in Exp. 2, which was set constant across blocks, which the negative stimuli interspersed; RBO-Exp3, Rule-Based Order in Exp. 3, sample for one participant, in which the presentation order was set constant across the learning blocks; the positive and negative stimuli were fully blocked in RBO-Exp3, with all the positive stimuli presented before all the negative stimuli; Stimulus numbers in the first column are indicated in Fig. 1.

Table 2
Mean inter-item similarity.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Rule	1.89	1.86	2.25	1.48
Sim.	2.23	2.25	2.36	-
dissim.	-	-	-	0.9

Note. The maximal inter-item similarity is 3 in Exp. 1-3, since two contiguous four-dimensional objects cannot have more than 3 features in common; The maximal inter-item similarity is 2 in Exp. 4, since two four-dimensional objects cannot have more than 2 features in common.

Method

Participants. The participants were 22 freshmen and sophomores at the Université de Franche-Comté (France) who received course credit in exchange for their participation.

Procedure. The procedure was similar to Experiment 1, except that, in the present study, constant presentation orders were used within subjects and between blocks. This procedure can presumably help the participants perceive subpatterns of responses (e.g., + - + -) that can be used to classify instances blindly. For instance, after noticing that a +-+ patterns occurs after a "large red hatched square", this pattern can be used as a cue to classify correctly three instances in a row. This is the reason why this condition was tested with a small sample of participants. The procedure balanced the number of participants between the two types of presentation orders.

Results

The mean inter-item similarity for each condition is given in Table 2. The learning curves in Figure 2 *B* show the influence of presentation order on proportion correct across blocks. Confirming order effects in Exp. 1, learning was faster in the rule-based order. A mixed-model analysis similar to the one conducted in Exp. 1 confirmed that both Presentation order ($F(1, 685.1) = 135$, $p < .001$) and Block number ($F(39, 41.0) = 28.9$, $p < .001$) had a significant effect on proportion correct, without interacting significantly. An estimate of the fixed effect of the rule-based presentation order showed a 95% confidence interval that the improvement seen in the rule-based condition was between .003 less and .13 more than for the similarity-based condition (the mean effect was equal to .063 across blocks).

When comparing Exp. 1 to Exp. 2, one can make the simple observation that the two curves of Exp. 2 visually separate sooner and better from the first few blocks and all along the remaining blocks. When the two experiments were added in the analysis as a fixed variable, we found a significant benefit of constant orders (Exp. 2) relative to randomized orders (Exp. 1). The variable Experiment (1 vs 2) effectively had a significant effect on performance ($F(1, 3229.2) = 197$, $p < .001$), and significantly interacted with presentation order ($F(1, 3229.2) = 42.9$, $p < .001$). This interaction showed that there was a differential effect of presentation order between the two experiments (a global difference of 2% across blocks in the first experiment, and 10% in the second experiment). The difference between the two experiments was estimated to be between .005 less to .19 more between the first and second experiment (with a 95% condence interval), with an average of .09 across blocks.

A follow-up questionnaire given at the end of the experiment indicated that 17 participants noticed that the presentation order was manipulated (but they were still unsure about the perfect regularity/circularity of the sequences), 5 more participants declared that they believed that the order was perfectly constant (3 in the similarity-based condition, and 2 in the rule-based condition), and only one did not notice any manipulation of the presentation order (in the rule-based condition).

Experiment 3

This experiment explores the effect of interleaving negative examples and positive examples. In Mathy and Feldman (2009) and in many earlier studies, while the order of the positives was the main manipulation of interest, negative examples were interspersed with positives in order to emulate an ordinary random presentation. But the effect of this interleaving, as opposed to a blocked presentation in which positives and negatives are segregated, is not known; here we explore this issue. One hypothesis is that when the stimuli are blocked, the perception of the commonalities within categories is favored in rule-based orders, hence resulting in faster learning. However, the perception of contrasts between categories might be enhanced because of the immediate juxtaposition of positives and negatives. As previously discussed, it is difficult to decide between the opposite effects of blocked *vs*. interspersed presentations without knowing exactly the type of category being studied (Carvalho & Goldstone, 2015). Here, because we use a difficult concept with highly discriminable categories (in which the stimuli are dissimilar both within and between categories), a blocked presentation should result in better performance using a rule-based presentation. One might also imagine that under an exemplar model the featural and temporal contiguity of exemplars within a category in a blocked presentation might reinforce intracategory associations, hence also resulting in faster learning.

Method

Participants. The participants were 46 freshmen and sophomores at the Université de Franche-Comté who received course credit in exchange for their participation.

Procedure. The task was similar to Exp. 2, except that the learning and categorization phases alternated, separated by a pause of 5 seconds. The presentation order was constant across the learning phases, as in Exp.2, but fully-blocked (the positive examples were always presented first, followed by the negative ones, Clapper & Bower, 1994, 2002). The presentation order (rule-based, $N = 23$ or similarity-based, $N = 23$) was again a between-subject manipulation.

During the learning phase, the stimuli were displayed first for one second. When the stimulus was presented, the category was displayed below the stimulus (i.e., a schoolbag or a trash can) and the corresponding category picture was also displayed for one second (for instance, the school bag was shown for one second, while the trash can was hidden for one second for a positive category). This was followed by a confirmation phase during which the participant had to press the response key corresponding to the category picture that had just been shown to them. After the key was pressed, feedback indicating a correct or incorrect classification was given at the bottom of the screen for two seconds. Our pretests showed that in this condition, the participants could not fail to correctly give any of the instructed responses. Therefore, the participants were expected to get 100% correct feedback during this phase. This method was used to make sure that the partici pants were following the learning phase actively and that they did not miss any of the instructed categories. The progress bar was hidden during the learning phase. Following the procedure of the previous experiments, the number of points that were accumulated on the progress bar was restored whenever a categorization phase began. The procedure for running the categorization phases were similar to those used in Exps. 1 and 2.

Results

The mean inter-item similarity for each condition is given in Table 2. One can note that blocking greatly increased the mean inter-item similarity. The learning curves in Figure 2 *C* show the influence of presentation order on learning. When comparing Exp. 2 to Exp. 3, we globally observed a reduction of about half the number of blocks required to reach the criterion in the third experiment (the criterion is basically reached at around 15 blocks in the rule-based order in the present experiment, instead of around 30 blocks in the second experiment). One could argue that because the participants benefited from a training block before the categorization block, the number of blocks during which the stimuli and their categories could be observed is 30 (instead of 15), which somewhat muddles the present effect of blocking. Nevertheless, we only count here the number of blocks in which the participants were tested.

A mixed-model analysis on proportion correct, similar to the one conducted for the two previous experiments but limited to the first 25 blocks (since most of the participants reached the learning criterion before the 25th block), showed that Presentation order ($F(1, 1034.4) = 32.8$, $p < .001$) and Block number ($F(24, 82.6) = 27.7$, $p < .001$) were both significant, but we observe no interaction effect ($F(24, 82.6) = .39$). Estimate of the fixed effect of presentation order showed that there was a 95% confidence that the improvement seen in the rule-based condition was between .04 less and .11 more than for the similarity-based condition (the mean effect was equal to .03).

When the second and third experiments were added in the analysis as a fixed variable, a significant effect of presentation order was again present ($F(1, 2058.7) = 51.4$, $p < .001$), as well as a significant effect of Block number ($F(24, 189.5) = 29.5$, $p < .001$). We observed a significant interaction between Experiments and Presentation order ($F(1, 2058.7) = 59.7$, $p < .001$) The global difference in performance between the two experiments was estimated to be between .06 less to .15 more between the second and third experiment (95% confidence interval), with an average of .05 across blocks. We also observed a significant interaction between Experiments and Block number, $F(1, 2058.7) = 125.3$, $p < .001$ that represent the steepest increase in proportion correct in the third experiment.

Results for Experiment 1-3 combined

We limited the mixed-model analysis of the three first experiments to the 25 first blocks. Table 3 shows the proportion correct for the three experiments, which increases with experiment number, and Table 4 indicates the result of the mixed-model analysis. Post-hoc analysis using Bonferroni correction showed that there was a systematic difference between the Experiments, with a mean difference in proportion correct across blocks of .07 between Exp. 1 and Exp. 2, .11 between Exp. 2 and Exp. 3, and .19 between Exp. 1 and Exp. 3. Most importantly, the mixed-model analysis showed a significant effect of Presentation order across experiments and a significant interaction between presentation order and Experiments indicating that the effect of the presentation orders (rule-based vs. similarity-based) was modified by their implementation within blocks (i.e., variable in Exp. 1, constant in Exp. 2, and blocked in Exp. 3)

Experiment 4

Method

Exp. 4 extends the previous studies to three 3D Boolean concepts, specifically types III, IV and V from the much-studied six types introduced by Shepard et al. (1961). The subjective difficulty of the six SHJ types follows the order I > II > (III , IV , V) > VI (Feldman, 2000; Nosofsky, Gluck, Palmeri, McKinley, & Gauthier, 1994; Shepard et al., 1961) with Types III, IV, V generally found to be approximately equal in difficulty. Types III, IV, V are illustrated in Fig. 1, using the three features of shape (circle vs square), color (green vs red) and filling (plain vs hatched). Size was set constant to 'Large'. In this experiment, we chose to compare the rule-based order with the dissimilarity-based order (between-subjects) in order to induce the most largest possible contrast between presentation orders (following the result of Mathy & Feldman, 2009). In a dissimilarity order, successive examples are chosen to be as *different* as possible, a condition found in Mathy and Feldman (2009) to be particularly detrimental to learning. The difficulty conveyed by this ordering presumably involves its tendency to shatter and thus obscure common properties of category items. Our goal here was to elicit a strong learning contrast to provide a challenge to the modeling.

Participants. The participants were 53 freshmen and sophomores at the Université de Franche-Comté who received course credit in exchange for their participation.

Procedure. Each participant was given the three concepts in random order and was assigned one type of presentation order, rule-based ($N = 24$) or dissimilarity-based ($N = 29$). The three SHJ Types are shown in an arbitrary rotation in Fig. 1, and rotations were randomly drawn for each concept and each participant, meaning that the diagnostic dimensions varied from one participant to another at random.

For each concept, the procedure was similar to Exp. 3 (learning was fully-blocked, and performance was measured during a random classification block). Maximal inter-item dissimilarity was only local in Mathy and Feldman (2009),

Table 3
Mean proportion correct across the 25 first blocks.

	Exp. 1	Exp. 2	Exp. 3
Rule	.69 (.005)	.82 (.008)	.90 (.006)
Sim.	.69 (.005)	.72 (.008)	.86 (.006)

Note. Standard errors are given in parentheses.

Table 4
Mixed model analysis for the first three experiments, presentation order (rule vs similarity), and block number (from 1st to 25th).

source	df num	df denom	F	Sig.
Intercept	1	3177.2	88423.5	.000
Exp.	2	3176.6	638.6	.000
Order	1	3177.2	82.7	.000
Block	24	243.5	51.9	.000
Exp. × Order	2	3176.6	33.1	.000
Exp. × Block	48	246.4	3.3	.000
Order × Block	24	243.5	.5	.973
Exp. × Order × Block	48	246.4	.5	.998

which means that it was only computed between two contiguous stimuli. To maximize the dissimilarity effect in the present experiment, the first stimulus in the dissimilarity-based order was chosen to offer the maximal overall distance within each category. For instance, in Type III, the red hatched square was chosen first, and was then followed by the red plain circle, the green hatched square, and the red hatched circle, totalizing a distance equal to $2 + 3 + 2 = 7$ (which is equal to a mean inter-item similarity equal to $(1 + 0 + 1)/3 = .7$). The rule-based order always matched the first chosen stimulus of the dissimilarity-based order, but then followed the simplest rules, that is, hatched squares and red circles in Type III, not(green hatched circle) and red plain square in Type IV, and not(green hatched circle) and green plain square in Type V. For example, when the order in Type III was red hatched square, green hatched square, red hatched circle, and red plain circle, the overall distance was equal to $1 + 2 + 1 = 4$ (which is equal to a mean inter-item similarity equal to $(2 + 1 + 2)/3 = 1.7$).

Results

The mean inter-item similarity for each condition is given in Table 2. As expected, the average inter-item similarity was the lowest for the dissimilarity-based condition. The learning curves in Figure 2 *D* show the influence of presentation order on proportion correct, with a positive effect of the rule-based order. A preliminary repeated measures ANOVA on proportion correct found no evidence of performance varying across Types III, VI and V, $F(2, 104) = 1.4$, $p = .25$. The data was therefore collapsed across Types to obtain a single measure of the proportion correct per block and per participant. A repeated measures ANOVA on proportion correct using Blocks (from 1st to 5th) as a within-subject factor and Presentation order as a between-subject factor showed a significant effect of Presentation order ($F(1, 51) = 5.0$, $p = .03$, $\eta_p^2 = .09$) and Block number ($F(4, 204) = 41.9$, $p < .001$, $\eta_p^2 = .45$), with no interaction effect between the two factors $F(4, 204) = .35$. Proportion correct was on average .06% higher in the rule-based condition (.89% instead of .83% in the dissimilarity condition). When an independent-samples t test was conducted on the first block, Presentation order already showed a significant effect on performance ($t(51) = 2.1$, $p = .041$, with a proportion correct on average .07% higher in the rule-based condition (.81% instead of .74% in the dissimilarity condition), which indicates that the beneficial effect of the rule-based presentation was extremely rapid.

Temporal-GCM

Exemplar models, at least in their traditional "batch" formulation, cannot predict or account for the effects of presentation order. Such models are generally insensitive to the order in which the stimuli are presented, as they lack any mechanism specifically designed to modulate the strength of the memory traces of the various exemplars. This does not mean that a standard exemplar model cannot fit *a posteriori* our data well for the various presentation orders in the different experiments. With its demonstrated capacity to fit data, a basic exemplar model such as Nosofsky's Generalized Con-

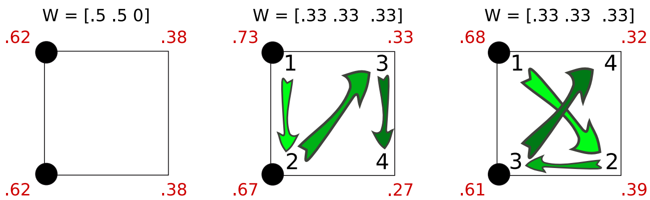

Figure 3. An example of how the temporal dimension affects the categorization probabilities in a temporal version of GCM. *Note*. *W* is a vector of dimension-salience weights. The two first values are the weights for the physical dimensions. The third value in the vector is the weight given to the temporal dimension. The left square shows the categorization probabilities of classifying a stimulus in the positive category (i.e., $p(A)$), when equivalent attention is allocated to both physical dimensions. The central square indicates the categorization probabilities of classifying a stimulus in the positive category, when the temporal dimension is given a .33 value, and when stimuli are presented by clusters in a fully-blocked manner (for every square, the value "1" represent the first displayed item, the value "2" represent the second displayed item, and so on). The right square shows the categorization probabilities of classifying each of the stimuli in the positive category when the temporal dimension is still given a .33 value, but when stimuli are presented in a more unstructured way (with the negative examples interspersed).

text Model (GCM) (Nosofsky, 1984, 1986; Nosofsky, Gluck, et al., 1994) might accommodate our data using suitable parameter settings. Such a fit would, though, shed little light on the mechanisms responsible for the observed presentation order effects.

Instead here we developed an exemplar model that does take the temporal dimension into account in simple way, including it in the computation of psychological distances, which we call the *Temporal Generalized Context Model* or TGCM. In TGCM, the perception of similarity between two stimuli can be influenced by their relative serial position, inducing a neighborhood structure in terms of both features and time. Similar ideas have been developed by Howard & Kahana, 2002, but the present model is much more minimalist. Time, in this account, simply becomes a new feature treated in a manner similar to the other features. Therefore, the model can capitalize on the temporal dimension to assess distinctiveness (Murdock, 1960) in the same way that serial position effects can be modeled as discrimination problems in serial or free recall tasks (Brown, Neath, & Chater, 2007).

In this section, we present TGCM and show that it can provide a reasonably good fit to the data presented above. TGCM is a very simple extension of the standard model, indeed so simple that it does really require any formal extension beyond the inclusion of the time dimension. In GCM, the distance function presented in Equation (1) can be used with $r = 1$ (a city-block metric adequate for separable dimensions), n the number of physical dimensions, and x_{ia} the value of stimulus i on dimension a. The distance function is augmented with two parameters: a scale parameter c reflecting discriminability in the psychological space, and n attention parameters modulating the weight of the n dimensions, $0 \le w_a \le 1$, and $\sum w_a = 1$ ($n - 1$ are free to vary).

$$d_{ij} = c[\sum_{a=1}^{n} w_a |x_{ia} - x_{ja}|^r]^{1/r} \tag{2}$$

Stimulus similarity decays exponentially with psychological distance (Nosofsky, 1986; Shepard, 1987):

$$\eta_{ij} = e^{-d_{ij}} \tag{3}$$

Given the total similarity of a stimulus i to all exemplars in categories X and Y, the probability of responding with category X is computed via Luce's choice rule[1]:

$$P(X|i) = \frac{\sum_{x \in X} \eta_{ix}}{\sum_{x \in X} \eta_{ix} + \sum_{y \in Y} \eta_{iy}} \tag{4}$$

In TGCM, the distance function presented in Equation (1) makes use of an extra attention weight to the temporal dimension. To simplify (following Brown et al., 2007, p. 544), r was set to 1 for both physical and temporal distances. We now detail why presentation orders can affect performance in this model. According to TGCM, the categorization process is influenced by the temporal contiguity between stimuli. Upon presentation of a stimulus, the psychological distance between the stimulus and the exemplars depends on both the physical dimensions and the temporal dimension. Given presentation order, TGCM can be primarily used to account for the distortion of the memory traces. Figure 3 illustrates how presentation orders can affect the pattern of probabilities when some attention is given to the temporal dimension (for instance, by setting the attention weight vector to $w = [.33\ .33\ .33]$, with the first two values corresponding to the two physical dimensions and the last value corresponding to the temporal dimension). When $w = [.5\ .5\ 0]$

[1] The choice rule can be augmented with γ, which is a response-scaling parameter that governs the extent to which responding is probabilistic versus deterministic (Ashby & Maddox, 1993; McKinley & Nosofsky, 1995; Navarro, 2007; Nosofsky & Zaki, 2002), a parameter which is useful to fit the data along the blocks, to better fit the data when performance may be close to the chance level (at the beginning of an experiment) and when performance is error less (sometimes by the end of an experiment). $P(X|i) = (\sum_{x \in X} \eta_{ix})^\gamma / [(\sum_{x \in X} \eta_{ix})^\gamma + (\sum_{y \in Y} \eta_{iy})^\gamma]$. Values of γ less than 1 reflect greater levels of guessing, whereas values above 1 makes the predicted probabilities more deterministic (close to 0 or 1). This parameter can be avoided when fitting the data by epochs. The parameter representing the bias for making category responses and the one controlling for the frequency of the stimuli were also considered minor in our study (respectively, because the number of positive and negative examples is balanced in $12_{4[8]}$).

(left square of Figure 3), the probabilities of classifying each of the positive examples as positive examples is $p = .62$, against $p = .38$ for the negative examples (no matter which presentation order is given because the attention weight of the temporal dimension is set to zero, and with c and γ both set to 1). When the presentation order is fully blocked (middle square of Figure 3) and when $w = [.33\ .33\ .33]$, the distinctiveness of the first positive exemplar presented increases contrary to the second, as regards to the negative examples ($p = .73$ for the first, against $p = .67$ for the second). The effect is reversed for the negative exemplars for which the first exemplar presented is given a higher probability of being categorized as a positive (because it is presented in the middle of a set of positive examples), contrary to the last negative exemplar presented ($p = .27$) that is isolated. Therefore, the first and last stimuli are better discriminated from the opposite category and better associated with their correct category (respectively $p = .73$ and $p = .27$). Overall, all examples are expected to be better categorized in this fully-blocked presentation order using $w = [.33\ .33\ .33]$, compared to the prediction that uses $w = [.5\ .5\ 0]$: effectively, the two probabilities for the positive examples($p = .73$ and $p = .67$) exceed $p = .62$; likewise for the negative exemplars for which the two probabilities ($p = .33$ and $p = .27$) are both inferior to $p = .38$. When the presentation is more disorganized (see the right square in Figure 3), the change in the predicted probabilities is more dramatic, with a higher probability that the first negative example will be misclassified ($p = .39$), and a lower probability that the second positive example ($p = .61$) will be correctly classified. We believe that this simple integration of the temporal dimension in the computation of the similarity matrices for each participant and each epoch can result in more precise predictions than the standard GCM.

To test TGCM, the fit measures were first computed for each experiment, each presentation order, each participant, and each epoch of five blocks. Following the simplifying assumption that the temporal distances were restricted to the block boundaries, the temporal distances were computed for each block. Because the distances were computed between one object and the preceding objects only within a block, the distance between each pair of objects was symmetrical. For example, for three objects presented in a block, the distances between the first and second objects, the first and third objects, and the second and third objects would be 1, 2, and 1 respectively. Therefore, we hypothesized that the participant would rely on temporal associations to guide learning, with associations being stronger in the forward than in the backward direction (Kahana, 1996; Kahana, Howard, & Polyn, 2008). This assumption is relevant because each presentation-order-manipulated block was separated by a categorization phase in Experiments 2 and 3, and because of a potential primacy effect in Experiment 1 (the fact that the first block began with a given set of objects that came back circularly). We assumed that these two different clues could help participants identify the temporal structures related to the beginning and end of the ordered blocks. This assumption is open to criticism for the first experiment, where the subjects had no clear idea where the blocks began after the first block, but we still wanted to systematically run the model in a similar fashion for all the data sets.

We scaled the temporal distances to equalize the maximal temporal distances with the maximal featural differences. (Specifically temporal distances were divided by 3.75, the value that most closely aligns with the scale of the 15/4 = 3.75, because the maximal temporal distance between two stimuli within a block is 15, i.e. the difference between 1, the first stimulus presented within a block, and 16, the last; while the greatest featural difference between stimuli is 4.)

This scaling method could be further amended by the attention weighting process described above. Note that because the temporal dimension is fully diagnostic of the categories in a fully-blocked presentation, we computed a mean temporal distance based on the distances in the presentation phase and those in the random categorization phase in the last two experiments.

Log-likelihood was used as a measure of goodness-of-fit.[2] The likelihood of the data was computed using the binomial distribution for each experiment, each participant, each epoch, and each parameter setting:

$$L = \prod_i \binom{F_i}{f_i} p_i^{f_i} (1 - p_i)^{F_i - f_i}, \tag{5}$$

with i the stimulus number, f_i the number of positive category responses for stimulus i, F_i the number of blocks in an epoch (i.e., the maximal number of positive category responses for stimulus i), and p_i the probability given by the model of giving a positive category response for stimulus i.

Because GCM is a restricted version of TGCM, a test of the difference of the goodness-of-fit between the models was done using the likelihood-ratio statistic (i.e., also called the difference of the -2 log-likelihood of two nested models):

$$\chi^2(df) = -2[\ln L(\text{restricted}) - \ln L(\text{general})] \tag{6}$$

The degrees of freedom are the number of parameters that are removed in the restricted model, compared to the general version. Here, $df = 1$ because the restricted GCM model does

[2] Computing the log likelihood simply amounts to replacing a computation based on a product (to compute the joint likelihood across participants and across epochs for all models and all set of parameters) by a computation based on a sum (for an introduction, see Lamberts, 1994). Effectively, it is more convenient and more accurate to compute the logarithm of the likelihood to avoid undesirable underflow which can result from constantly multiplying values between 0 and 1.

not integrate the temporal dimension present in the more general TGCM model.

First, both GCM and TGCM were tested using a single sensitivity value and a fixed gamma value (both equal to 1). The objective was to show the role of the temporal dimension for the simplest version of the model. The parameter of interest in this first investigation was w. We used all of the 126 possible weight patterns that could be generated using .2 steps in 5D (e.g., [1 0 0 0 0], [.8 .2 0 0 0], ..., etc.; using .1 steps, the set of 1001 different weight combinations significantly increased the search time of the model-fitting process) and all of the 56 possible weight patterns that could be generated using .25 steps in 4D (e.g., [1 0 0 0], [.75 .25 0 0], ...). All of these possible combinations of weight were tested for each epoch and each participant. Then, we computed which sequence of patterns of weights gave the best joint likelihood across the epochs for both models (for a given participant, the best weight patterns could therefore be different between epochs).

For the first three experiments, we conducted a stimulus-by-stimulus analysis of the responses across participants, epochs, and experiments, totalizing 16672 data points (68 participants $\times$ 8 epochs $\times$ 16 stimuli + 22 participants $\times$ 8 epochs $\times$ 16 stimuli + 46 participants $\times$ 7 epochs $\times$ 16 stimuli =16672). The maximum log likelihood for the three experiments were respectively -14956, -4412, -3874 for TGCM and -15577, -4573, -5311 for GCM. The respective summed maximum log likelihood across the experiments was -23242 and -25461 (see Table 5). Based on these sum log-likelihood across experiments, we obtained a series of χ^2 values across experiments that always exceeded the critical 3.84 value (for instance, $\chi^2 = -2[2649 - 2833] = 368$ for the first experiment). This result demonstrates that exemplar models can account for a significantly greater portion of the variance when taking into account simple temporal effects. These summed maximum log likelihood (-23242 and -25461) lead to the respective 46940 and 51377 AIC values (smaller is better) using 4 and 3 free parameters respectively since the sensitivity parameter was maintained constant (the BIC values were respectively 46971 and 51400).

We conducted a separate analysis for Exp. 4 and found a maximum log likelihood estimation of -2649 for TGCM and -2833 for GCM, again exceeding the critical χ^2 value and demonstrating the superiority of TGCM over GCM.

Table 5 shows the respective summed maximum log likelihood (-14254 vs -18052) for TGCM and GCM using the best attention weights (again, using the 126 different weight patterns in 5D) and the best sensitivity values c (when c was allowed to vary from 1 to 5 using increments of .5). The result lead to the respective 46940 and 51377 AIC values (smaller is better) using 5 and 4 free parameters respectively since the sensitivity parameter was this time allowed to vary (the BIC values were respectively 46971 and 51400). Figure 4 plots the simulation and observed proportion correct for the three experiments and the two versions of the model. The comparison of the first subplot and the second subplot shows the better fit of the data by TGCM. We again conducted a separate analysis for Exp. 4 (using the 56 weight patterns in 4D and c varying from 1 to 5 using increments of .5) and found a maximum log likelihood estimation of -1383 for TGCM and -1626 for GCM, which again allowed to exceed the critical χ^2 value.

General Discussion

Previous studies on inductive learning have rarely considered the order in which examples are actually encountered (Komatsu, 1992; Kruschke, 2005; Murphy, 2002). Our research addresses this question, with particular focus on how the effectiveness of presentation orders relates to the nature of the category be learned. This issue was first raised in Mathy and Feldman (2009), who reported an experiment in which a rule-based order facilitated greater category learning than an order that simply maximizes inter-item similarity (Elio & Anderson, 1981, 1984; Medin & Bettger, 1994). Following a rule-plus-exception approach (Nosofsky, Palmeri, & McKinley, 1994; Palmeri & Nosofsky, 1995), a rule-based presentation order is one that derives from the logical structure of the training examples: objects that are within a rule are presented adjacently and randomly in the presentation sequence, and training then moves on to "exceptions" until all the objects have been presented. A similarity-based order, by contrast, simply maximizes the similarity between consecutive items. The above studies included several additional manipulations, including constant presentation orders across blocks for each participant (Exp. 2, and Exp. 3), as opposed to the variable presentation orders across blocks used in Exp. 1 and in Mathy and Feldman (2009). We also made use of fully-blocked presentation orders so that the positive examples were separated from the negative examples during the training phases (Exp. 3 and Exp. 4). While Exp. 1-3 all used the same conceptual structure, Exp. 4 used a set of 3D concept types originally studied by Shepard et al. (1961).

Our main results include the systematic learning advantage for rule-based presentation orders in all four experiments (confirming Mathy & Feldman, 2009); a positive effect of constant orders (when contrasting the results of Exp. 2 to those of Exp. 1); and a positive effect of fully-blocked orders (when contrasting the results of Exp. 3 to those of Exp. 2). The blocked presentation reduces approximately by half the number of categorization blocks required to reach almost perfect performance. One limitation of the results in Exp. 2 is that constant orders can unfortunately allow the participants to memorize constant patterns of category responses (e.g., +-+, from a given starting point). Further investigation would therefore be required to better assess the effect of nonblocked constant orders, using for example a training

Table 5
Maximum log likelihood estimation, Akaike Information, and BIC values, for TGCM and GCM, as a function of Exp1, Exp. 2, Exp. 3.

	TGCM	GCM
A. Simulation with c = 1		
	TGCM	GCM
k	4	3
$\hat{\theta}_{mle} Exp.1$	-14956	-15577
$\hat{\theta}_{mle} Exp.2$	-4412	-4573
$\hat{\theta}_{mle} Exp.3$	-3874	-5311
$\hat{\theta}_{mle} Total$	–23242	-25461
AIC	46940	51377
BIC (N = 16672)	46971	51400
Simulation with c = [1 : .5 : 5]		
	TGCM	GCM
k	5	4
$\hat{\theta}_{mle} Exp.1$	-9487	-11207
$\hat{\theta}_{mle} Exp.2$	-2587	-3212
$\hat{\theta}_{mle} Exp.3$	-2180	-3633
$\hat{\theta}_{mle} Total$	-14254	-18052
AIC	28967	36562
BIC (N = 16672)	29006	36593

Note. k, number of parameters free to vary; $\hat{\theta}_{mle}$, maximum log likelihood estimation; AIC, Akaike Information Criterion (smaller is better).; BIC, Bayesian Information Criterion (smaller is better). Probabilities $p(A)$ predicted by TGCM used 126 attention different weight patterns generated by considering all possible combinations of weights that could be built using .2 increments (e.g., [1 0 0 0 0], [.8 .2 0 0 0], ..., [.6 .2 .2 0 0]. The sensitivity parameter was set constant in subtable A (c = 1) or could vary from 1 to 5 using .5 increments in subtable B. Probabilities $p(A)$ predicted by GCM used the same combinations of parameters except that the weight patterns for which attention to the temporal dimension was superior to zero were removed from the list of weigth patterns. Maximum log likelihood estimation was estimated in TGCM and GCM for each participant and each epoch using the best combination of attention weights and the best c. When c is constant, k is equal to 4 in TGCM because one of the five attentionnal weights cannot vary freely; k is equal to 3 in GCM when c is constant, because the attention weight to the temporal dimension is set to zero.

phase and a categorization phase as in Exp. 3. The results of Exp. 2 are still important to show that potential acceleration of learning is possible when repeating identical categorization blocks, without confounding the acceleration of learning with the use of the training phase used in Exp. 3. A second limitation is that constant blocked-orders effectively double the number of stimuli that are presented to the participants because of the use of the training phase that precedes the categorization phase. The effect of blocking would have been more convincing if, for instance, the number of categorization blocks had reduced by more than a half. The effects of the constant factor and the Blocked factor are therefore not totally conclusive, though sufficiently rich to allow a contrast between our Temporal GCM and GCM. Conversely, without additional data the effect of blocking by itself since our conditions did not included a non-constant-blocked procedure. Again, as mentioned above, the total number of potentially interesting combinations of conditions is too large to study comprehensively, so our focus was on a pulling out a particularly revealing interesting set of conditions suitable for modeling.

We make one final observation about the superiority of blocked over interleaved orders demonstrated in Exp. 3. This effect was observed consistently in both rule-based and similarity-based orders. This finding relates to earlier observations that participants have difficulty learning concepts from negative instances, when intertwined with positive instances, even though each category transmits the same amount of information (Hovland & Weiss, 1953), and suggests the interpolation of the negative instances has a detrimental effect on learning because it distracts observers from the common aspects of the positive category.

The effect of blocked orders itself, or more broadly the effect of grouping of positive instances, has previously been observed in the learning of paired-associate lists (Gagné, 1950), in supervised concept learning (Kurtz & Hovland, 1956; Goldstone, 1996), in unsupervised concept learning (Clapper & Bower, 1994, 2002; Zeithamova & Maddox,

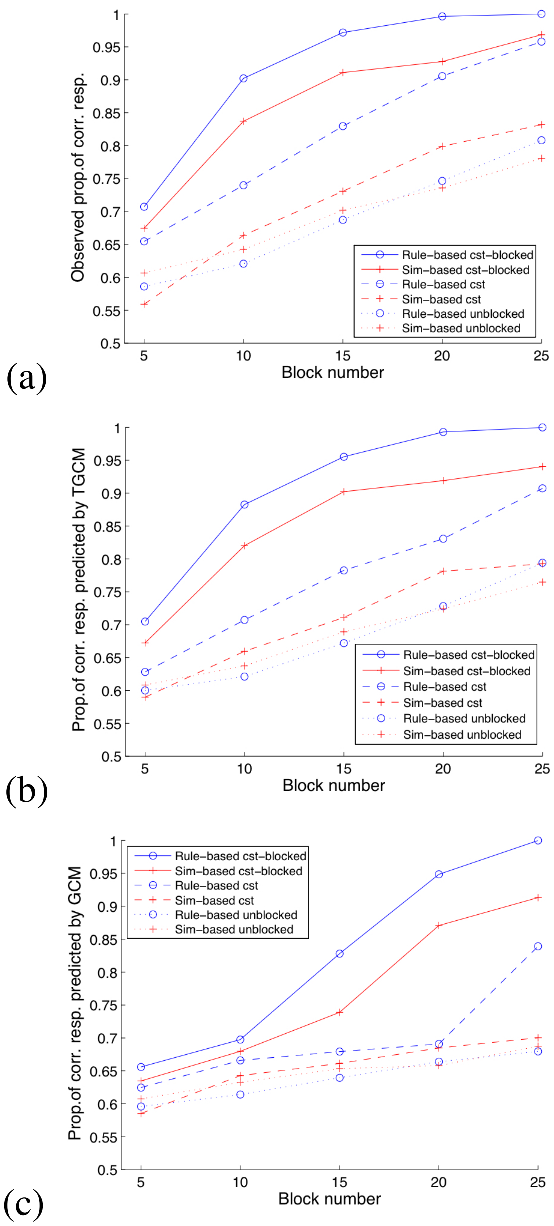

Figure 4. The continuous, discontinuous, and dotted curves correspond to Exp. 3, Exp. 2 and Exp. 1 respectively. The discontinuous curves correspond to the unblocked conditions of Exp. 2. (a) Proportion correct observed (b) Proportion correct predicted by TGCM (c) Proportion correct predicted by GCM.

2009), in incidental concept learning (Wattenmaker, 1993), and even in clustering (Gmeindl, Walsh, & Courtney, 2011). However, the fact that blocking categories benefits learning is somewhat at odds with the nearly ubiquitous finding of benefits connected to distributed as opposed to massed practice (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Spacing can appear be beneficial in some contexts, even though massing apparently creates a sense of fluent learning (Kornell & Bjork, 2008; Kornell et al., 2010; Wahlheim et al., 2011). Overall, our study suggests that isolating items connected with a principal regularity (i.e. separating the exceptions from the rules) benefits the identification of different clusters (it is easier for individuals to form a new clause when examples of different clusters are clearly separated); while blocking can benefit the formation of one cluster (since it is easier for individuals to formulate a clause based on a group of contiguous stimuli). In the context of an exemplar model, isolation and blocking can only benefit learning if one assumes that these manipulations limit the number of inter-stimulus associations, including temporal ones. Kornell and Bjork (2008) also noted that when discrimination is easy (Kurtz & Hovland, 1956), massing may be advantageous, but if discrimination is difficult, as in their experiments, spacing might be more effective. This observation does not apply to conditions in our experiments where the discrimination of examples is particularly difficult. For example, in concept $12_{4[8]}$, the positive and negative clusters are very intertwined, so a positive member (e.g., 0100) can be equally close to another member of the first positive cluster (e.g., 0000), a positive member of the second cluster (e.g., 1100), or a negative example (e.g., 0110).

Overall, our results point to an intricate relationship between time and categorization that adds to previous studies on learners' ability to restructure their knowledge in light of additional information (Lewandowsky, Kalish, & Griffiths, 2000), in this case temporal information. (Lee et al., 1988; Spiering & Ashby, 2008). Our manipulated orders (rule-based, similarity-based, dissimilarity-based) have a substantial effect on the way concepts are learned. Moreover, we developed a model, TGCM, that explains these effects in a fairly simple way, by extending the feature space to include time (or, more properly, serial order). To our knowledge, no other exemplar models can account for our results. This model describes how temporal proximity between stimuli "distorts" the psychological space generated by the examples. More negatively, our results suggest that the neglect of serial order effects in standard models is unwarranted. In most naturalistic conditions, examples are after all encountered in some definite order, and category learning evolves progressively rather than waiting for the end of relevant examples. The result, in real learning, is inevitably that the category learned is to some extent a function of the examples observed so far, and hence progresses differently depending on the order in which they are observed. From this perpsective it should be no suprise that the effect of presentation orders is so pronounced.

Admittedly though our study did not explicitly include other important classes of models, such as prototype models (Ashby & Maddox, 1993; Minda & Smith, 2001, 2002; Nosofsky & Zaki, 2002; Osherson & Smith, 1981; J. D. Smith & Minda, 1998; Zaki, Nosofsky, Stanton, & Cohen, 2003) and hybrid models (Ashby et al., 1998; Anderson & Betz, 2001; Erickson & Kruschke, 1998; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Nosofsky, Palmeri, & McKinley, 1994; Rosseel, 2002; E. E. Smith & Sloman, 1994; Vandierendonck, 1995; Vanpaemel & Storms, 2008), so these conclusions are necessarily provisional. With that in mind, though, the relatively weak benefit of similarity-based orders might be attributed to either the overly specific hypotheses they tend to induce in the participants' minds, a phenomenon that might be explained by a progressive model such as SUSTAIN which is intrinsically incremental (Love et al., 2004), or by an exemplar model if time is incorporated as a feature as in TGCM. The superiority of rule-based orders can also be viewed as relating to the benefit of isolating "exceptions"—essentially the von Restorff effect (Restorff, 1933) transposed to categorization. Similarly, the random ordering of items within clusters characteristic of rule-based orders aids in the discovery of diagnostic features, which connects to findings from the 1950s about the how the serial order of examples aids in the elimination of irrelevant features (Bruner, Goodnow, & Austin, 1956; Hovland & Weiss, 1953).

Our results also confirm some observations that have been made about the relationship between temporal distinctiveness and memory retrieval. For example, (Kahana, 1996) reanalyzed a number of classic free-recall studies and showed that temporal contiguity clearly determines some associative mechanisms, leading to a process of episodic clustering: regardless of their degree of semantic association, neighboring items studied in list positions tend to be reported successively and more rapidly during the recall period. Accordingly, rule learning can be considered as a discrimination problem: temporally separated items are more easily discriminated, while temporally contiguous items are less so. Brown et al. (2007) have also shown such a dependency between time and discriminability in their model of memory retrieval, incorporating temporal discrimination as a core principle to account for the fact that forgetting is due to reduced local distinctiveness.

Remaining issues

As explained above, our comparison of order types and blocking conditions is not comprehensive, and a number of empirical questions remain unanswered. For example, the relative efficacy or dissimilarity ordered and random orders is unknown. Perhaps more importantly, a larger and more comprehensive set of concepts should be tested, including both concepts that tend to induce rule-learning strategies as well as those that favor information-integration or prototype-distortion strategies, which might involve respectively the formation of exemplars and prototypes (see Ashby & Ell, 2001). The effect of blocking is also poorly understood, as no study using a non-constant-blocked procedure has yet been carried out; likewise the effect of constant orders, in the absence of a study including nonblocked constant orders using a training phase and a categorization phase. Finally, other models capable of explaining the known effects of presentation order should be developed to offer an alternative to TGCM.

Conclusion

Our results confirm the systematic benefit of rule-based orders on learning, as well as a smaller benefit of fully-blocked orders and a number of other robust order effects. More broadly, these results collectively suggest that the sequence in which items are presented distorts the psychological stimulus space in a way that has very concrete effects on learning. The interaction between presentation orders and learning is a complex issue that depends on whether the chosen presentation order aids or inhibits the progressively developing concept representation, which in turn depends on the nature of the concept to be learned. A more complete understanding of these issues can probably only be achieved via a more comprehensive investigation of a range of concept types and presentation orders. However the good data fits provided by the TGCM model suggest that great progress can be made by simply incorporating temporal aspects of the stimuli into existing models in a relatively straightforward way. Clearly, given the large and robust effects observed, and the obvious application to real-life learning environments, the issue of presentation order deserves further study.

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3-19.

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

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442-481.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204-210.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(372-400).

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013a). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Memory & cognition*, *41*, 392-402.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013b). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Memory & cognition*, *41*, 392-402.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539-576.

Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.

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., & Goldstone, R. L. (2012). Category structure modulates interleaving and blocking advantage in inductive category acquisition. In *Proceedings of the thirty-fourth annual conference of the cognitive science society* (pp. 186–191). Austin, TX: Cognitive Science Society.

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

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

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380.

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

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

Clapper, J. P., & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 443-460.

Clapper, J. P., & Bower, G. H. (2002). Adaptative categorization in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 908-923.

Elio, R., & Anderson, J. R. (1981). 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.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107-140.

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.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological review*, *119*, 223-271.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630-633.

Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology*, *47*, 75-89.

Gagné, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology*, *40*, 61-73.

Garner, W. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gmeindl, L., Walsh, M., & Courtney, S. (2011). Binding serial order to representations in working memory: a spatial/verbal dissociation. *Memory & Cognition*, *39*, 37-46.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178-200.

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

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*, 108-154.

Hahn, U., & Chater, N. (1998). Similarity and rules: distinct? exhaustive? empirically distinguishable? *Cognition*, *65*, 197-230.

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.

Homa, D., Rhoads, D., & Chambliss, D. . (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 11–23.

Hovland, C. I., & Weiss, W. (1953). Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology*, *45*, 175-182.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269 - 299.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(103-109).

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger (Ed.), *Cognitive psychology of memory. vol. 2 of learning and memory: A comprehensive reference, 4 vols (j. byrne, editor-in-chief).* Oxford: Elsevier.

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

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*, 500-526.

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

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

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*, 225-247.

Kruschke, J. K. (2005). Category learning. In K. Lamberts

& R. L. Goldstone (Eds.), *The handbook of cognition, ch. 7* (p. 183-201). London: Sage.

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology*, *51*, 239-243.

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.

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1003-1021.

Lee, E. S., MacGregor, J. N., Bavelas, A., Mirlin, L., Lam, N., & Morrison, I. (1988). The effects of error transformations on classification performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 66 - 74.

Lewandowsky, S., Kalish, M., & Griffiths, T. L. (2000). Competing strategies in categorization: Expendiency and resistance to knowledge restructuring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1666-1684.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309-332.

Markant, D., & Gureckis, T. (2014). Is it better to select or to receive? learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*, 94–122.

Mathy, F. (2010). The long term effect of relational information in type VI concepts. *European Journal of Cognitive Psychology*, *22*, 360-390.

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

Mathy, F., & Feldman, J. (submitted). Presentation order effects on category generalization.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: stable tendencies for learning exemplars versus abstracting rules. *Journal of experimental psychology. General*, *143*, 668–693.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(128-148).

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

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. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology*, *27*, 775–799.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 275-292.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review*, *67*, 16- 31.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Navarro, D. J. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, *51*, 85-98.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Gauthier, P. (1994). Comparing models of rules-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*, 352-369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53-79.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(28), 924-940.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*, 35-58.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 548-568.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences*, *28*, 1-14.

Restorff, H. von. (1933). Uber die wirkung von bereichsbildungen im spurenfeld [on the influence of the segregation in the trace field]. *Psychologische Forschung*, *18*, 299-34.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Simlilarity and analogical reasoning (pp. 21-59).* Cambridge, MA: Cambridge University Press.

Rosch, E., & Mervis, C. (1975). Family resemblances : studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, *46*, 178-210.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of experimental psychology. General*, *133*, 63–82.

Sakamoto, Y., Jones, M., & Love, B. C. (2008). Putting the psychology back into psychological models: Mechanistic versus rational approaches. *Memory & Cognition*, *36*(1057-1065).

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, 13, whole No. 517.

Skorstad, J., Gentner, D., & Medin, D. L. (1988). Abstraction processes during concept learning: a structural view. In *Proceedings of the 10th Annual Conference of the Cognitive Science Society* (p. 419-425). Hillsdale, NJ: Lawrence Erbaum Associates.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*, 167-196.

Smith, E. E., & Sloman, S. A. (1994). Similarity- vs. rule-based categorization. *Memory & Cognition*, *22*(4), 377–386.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science*, *19*, 1169-1177.

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

Thibaut, J. P., Dupont, M., & Anselme, P. (2002). Dissociations between categorization and similarity judgments as a result of learning feature distributions. *Memory & Cognition*, *30*, 647-656.

Thibaut, J.-P., & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1403 - 1415.

Vandierendonck, A. (1995). A parallel rule acti vati on and rule s ynt he s i s mo de l a parallel rule activation and rule synthesis model for generalization in category learning. *Psychonomic Bulletin & Review*, *2*, 442-459.

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: the varying abstraction model of categorization. *Psychonomic Bulletin & Review*, *15*, 732-749.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*, 750-763.

Wattenmaker, W. D. (1993). Incidental concept learning, feature frequency, and correlated properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 203-222.

Weitnauer, E., Carvalho, P. F., Goldstone, R. L., & Ritter, H. (2013). Grouping by similarity helps concept learning. In *Proceedings of the thirty-fifth annual conference of the cognitive science society.* (p. 3747-3752). Berlin, Germany.

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

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. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 731-741.