

A rule-based presentation order facilitates category learning

Fabien Mathy

Université de Franche-Comté – France

Jacob Feldman

Rutgers University – New Brunswick, USA

We investigated the mechanisms by which concepts are learned from examples by manipulating the *presentation order* in which the examples are presented to subjects. We introduce the idea of a *rule-based* presentation order, which is a sequence that respects the internal organization of the examples within a category. We find that such an order substantially facilitates learning as compared with previously known beneficial orders such as a similarity-based order. We discuss this result in light of the central distinction between rule-based and similarity-based learning models.

A number of studies have investigated whether category learning is influenced by the order in which examples are presented. Elio and Anderson (1981) found that categories are learned faster when training is blocked into groups of mutually similar examples (see also Elio & Anderson, 1984). More recently, Medin and Bettger (1994) demonstrated a strong learning advantage when training objects are presented in an order that tends to maximize similarity between successive examples. Other studies, such as the those of Clapper and Bower (1994) and Goldstone (1996), have focused on the effect of alternation of contrasting categories. Presentation order effects are especially interesting in light of categorization models that emphasize incremental learning from trial to trial. For example, Sakamoto, Jones, and Love (2008) showed that order can affect the incremental update of both category means and variances (see also Love, Medin, & Gureckis, 2004). Incremental learning models are naturally susceptible to order effects, while other models may be less so, so the manipulation of presentation order is a potentially useful tool for studying the mechanisms of learning.

However, previous studies of presentation order were limited in that they used orders based on simple similarity, for example maximizing or minimizing the similarity between adjacent training examples. Here we explore a type of presentation order that depends in a more “structured” way on the nature of the category to be learned. We introduce the notion of a *rule-based* presentation order, which is one that derives from the internal structure of the training examples. In our rule-based order, objects that are “within a rule”—that is,

that obey the same structured subclass within the category—are presented adjacently in the presentation sequence. Training then moves on to another subclass, and so forth until all objects have been presented. (Negative instances are randomly interspersed among the positives; only the order of the positives is manipulated.) Below, we compare subjects’ performance with such an order to the *similarity-based* order found to be advantageous in earlier studies. For comparison we also include a *dissimilarity-based* order, previously found to be disadvantageous. We hypothesize that the rule-based order would facilitate learning, particularly in highly structured concepts (i.e. those containing more clusters), by aiding the subject in mentally organizing what would otherwise appear heterogeneous or chaotic.

Method

Participants

The subjects were 96 Rutgers University students who received course credit in exchange for their participation.

Procedure

Tasks were computer-driven. Participants learned to sort stimulus objects using two keys, with successful learning encouraged by means of a progress bar. Stimulus objects were presented one at a time in the upper part of the computer screen. After each response, feedback indicating a correct or incorrect classification was provided at the bottom of the screen for two seconds. Subjects learned a simple concept in two dimensions as a short warm-up session. Then each subject was asked to learn the two chosen concepts (details below). The order of the two concepts was counterbalanced between subjects. For each subject, a single presentation order was randomly chosen and applied to the two concepts. The three different presentation orders are described in more detail below.

Each correct response scored them one point in a progress bar. The point was represented by an empty box that was filled in when subjects gave a correct response. To regulate the learning process, each response had to be given in

This research was supported in parts by a postdoctoral research grant from the Fyssen Foundation to Fabien Mathy and by NSF 0339062 to Jacob Feldman. We are grateful to Jennifer Biddick, Jonathan Geis, and Jing Tien for assistance in data collection, and to Cordelia Aitkin, Erica Briscoe, and David Fass for helpful discussions.

Correspondence concerning this article should be addressed to Fabien Mathy, 30-32 rue Mégevand, 25030 Besançon Cedex, France. or by e-mail at fabien.mathy@univ-fcomte.fr.

less than 8 seconds (making a maximum of 10 seconds between two stimuli when the subjects got a “Too late” message feedback lasting itself two seconds). If the response was given too late, participants would lose 3 points on the progress bar. The number of points in the progress bar dedicated to learning was 4×2^D (D = number of dimensions, 4 in our study). This criterion was identical to the one used by Shepard, Hovland, and Jenkins (1961) in their first experiment. Consequently, subjects had to correctly classify stimuli on four consecutive blocks of 2^D stimuli.¹

Choice of concepts studied

Participants were each given two concepts to learn, each defined over four dimensions (shown schematically in Fig. 1). We used four-dimensional concepts so that the number of objects to be classified ($2^4 = 16$) would be large enough to bring out any effects of our manipulations of presentation order, but small enough to be manageable in a single experimental session. We chose to focus on types $1_{4[8]}$ and $12_{4[8]}$ (Fig. 1) of the typology of Feldman (2003) (an extension of those in Shepard et al., 1961 and Feldman, 2000). In this notation, the [8] extension means that there are 8 positive examples in the concept, 4 means that the concept is based on four dimensions, and the 1 and 12 are arbitrary labels which identify these concepts from among the 72 other concepts available.²

The two concepts are defined up to isomorphism respectively by the formulae

$$1_{4[8]} \cong d',$$

$$12_{4[8]} \cong a'(bc)' + ad'(bc' + b'c).$$

Here we use a standard notation in which a' refers to negation of feature or clause a , ab refers to the conjunction of a and b , and $a + b$ to their disjunction (equivalent to but more concise and readable than the \wedge , \vee , \neg notation often used). In Fig. 1 the concepts are shown in an arbitrary rotation and permutation of features; as explained below this mapping was randomized in the experiment. The concepts are also encoded and represented linearly in Table 1.

The symbol \cong in the equations above indicates congruence or structural isomorphism up to this arbitrary mapping. For example the concepts a , a' , b , ..., or d' are all congruent because they are equivalent after relabeling of the features; in each case exactly one value of one feature defines the concept. Concept $1_{4[8]}$, because it can be expressed by a single literal, has complexity 1 and is thus the simplest concept in the $4[8]$ family, equivalent to “affirmation” (assertion of the presence of a single feature) in the classical literature (the four-dimensional analog of Type I from Shepard et al., 1961).

The second concept, $12_{4[8]}$, has complexity 9 literals and is thus of moderate complexity relative to others in the $4[8]$

family. (Complexity of concepts in this family ranges from 1 to 22 literals.) We chose this concept for several reasons. First, we wanted a moderately complex concept so that the entire learning procedure could be completed by most subjects in about an hour. Among concepts of moderate complexity, we chose $12_{4[8]}$ because its positive examples can be grouped fairly naturally into *subcategories* or clusters (labeled Clusters 1, 2 and 3 in Fig. 1), allowing us to investigate the interaction between presentation order and such internal substructure. As can be seen in the figure, Cluster 1 comprises six of the concept’s eight members, corresponding to the first disjunctive clause ($a'(bc)'$) in the concept’s compressed formula. Thus these six objects collectively receive an extremely compact expression—a clause of only three literals (which can be translated into a verbal expression such as “all a' except bc ”). By contrast, Clusters 2 and 3 consist of only one object each, each requiring four literals to specify just by themselves (respectively $abc'd'$ and $ab'cd'$, corresponding to the expansion of the second clause in the formula). Thus Cluster 1 plays the role of a salient “rule”, while Clusters 2 and 3 play the role of “exceptions”.

Our presentation order is based on the presupposition that subjects will cluster the members of concept $12_{4[8]}$ in the manner given above. This a reasonable assumption, in part because this clustering corresponds to a highly compressed Boolean form, consistent with the minimization of Boolean complexity (Feldman, 2000). Naturally, though, this concept (like any other) admits other interpretations or subclusterings, and we have no way of confirming that our subjects mentally organized it in the way we expected (other than the fact that the presentation order based on this decomposition *did* in fact benefit learning, as will be seen below). However, any alternative subclustering that subjects might have drawn would simply add noise to our analysis, working against our hypothesis, so our assumption is conservative.

Stimuli

Stimulus objects varied along four binary, separable dimensions (shape, color, size, and filling texture). For each concept (designation of 8 positive and 8 negative objects), assignment of abstract conceptual structures to physical features was randomized, i.e. the abstract features a , b , c , d were randomly permuted before being realized as physical features. Also for each concept, the choice of two values for each feature was chosen randomly from a sometimes longer

¹ Many studies require only 75% or 80%, but such a low criterion would defeat the goals of our study, because it would make it possible to learn only the “rule-like” examples and completely avoid the “exceptions.”

² In four Boolean dimensions, there are 74 qualitatively different types of concepts with 8 positive examples, giving a classification analogous to (though more complex than) that of Shepard et al. (1961).

list (shape = triangle, square, or circle; color = blue, pink, red, or green; filling = hatched or grilled; size = small or big). Each of the 16 combinations of values formed a single unified object (e.g., a small hatched red square, or a big grilled blue circle) to avoid numerical or spatial biases when displaying stimuli.

Ordering of stimuli

The main manipulation in the experiment was the choice of presentation order of objects within each concept. Presentation order was a between-subject manipulation. Again, one presentation order was randomly chosen for a given subject and then applied to the two concepts. We used three orders: a *rule-based order*, a *similarity order*, and a *dissimilarity order*. As discussed above, concept $12_{4[8]}$ consists of a single relatively coherent cluster of six objects and two “clusters” of one object each. By contrast, concept $1_{4[8]}$ consists of a single homogeneous cluster.

In each of the three order types, objects were drawn without replacement in each block of 16; that is, each block of 16 consisted of a complete permutation of all objects. The negative examples were randomly interspersed among the positives (in random order) in order to avoid long uninterrupted sequences of positives or negatives, even though this presumably made it more difficult for the subjects to benefit from the presentation orders. Again, only the order of the positive examples was manipulated, taking one of the orders detailed below, while the order of the negatives was always random (and thus different from block to block and subject to subject).

In the **rule-based order**, objects were drawn randomly from within the largest cluster (in $1_{4[8]}$, the entire concept; in $12_{4[8]}$, Cluster 1) until all 8 ($1_{4[8]}$) or 6 ($12_{4[8]}$) had been presented. In concept $1_{4[8]}$ this would exhaust the entire concept, while in concept $12_{4[8]}$, this would be followed by the objects in Cluster 2 and Cluster 3 (in random order). Thus in the rule-based order all members obeying a common rule were presented together, in random order but separated from exceptional members.³

In the **similarity 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 concept was exhausted.⁴ Ties were resolved randomly. Dissimilarity between two stimuli i and j is given by the Minkowski metric

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

where x_{ia} is the value of stimulus i along dimension a . We used a city-block metric appropriate for separable dimensions used in this study ($r = 1$). In general, this similarity ordering does not respect the cluster boundaries in force in

the rule-based order, as similarity steps routinely cross in and out of clusters in $12_{4[8]}$. The similarity order also differs from the rule-based order in that (aside from ties) the steps are not random.

In the **dissimilarity order**, objects were drawn exactly as in the similarity order except with similarities minimized instead of maximized. That is, each object would be followed by another object as distant as possible from it in the space.

In all three orders, each new block of 16 was newly randomized (the positive instances were randomly drawn but constrained to obey the desired order, and the negative instances were randomly interspersed) so subjects rarely saw consistent specific sequences of objects between blocks.

Comparison to procedures used in other studies

In Elio and Anderson’s (1981) similarity-based order, the presentation order increased inter-item similarity relative to a random sequence, but did not maximize it. Our similarity-based order is based on the Minkowsky metric, with no distinction of any sort between examples except their similarity. Items are simply chosen so that they maximize (or in the dissimilarity order, minimize) the similarity to the previous item. In this sense our similarity-based order is more extreme than Elio and Anderson’s, but also more varied than the one used by Medin and Bettger (1994), who used a single fixed similarity-based order and dissimilarity-based order in each of their experiments. Our procedure produces a locally maximal inter-item similarity in the similarity-based order, and a minimal inter-item similarity in the dissimilarity-based order. (Numbers are given in the results section.) In comparison, the inter-item similarity in the rule-based order is moderate. Some presentation order samples are given in Table 1. Finally, note also that because subjects are not aware of where the blocks begin, subjects might be more sensitive to the isolation of the stimuli belonging to different clusters in the rule-based ordering than to the strict position of the clusters in the blocks (related to the von Restorff isolation effect, transposed to categorization in Sakamoto & Love, 2004).

³ The term “rule-based” refers to the fact that this order respects a “rule plus exception” organization (Nosofsky, Palmeri, & McKinley, 1994). The organization more generally relates to disjunctive normal form (DNF), in which each term indicates a conjunction of features. Some disjunctive terms cover many cases (major rules); others cover fewer (minor rules, or major classes of exceptions); and still others cover only case one each (exceptions). An example is a rule like “birds = (animals that fly) or (ostrich-like animals— which possess feathers but do not fly) or (kiwi—also a flightless bird but not like an ostrich)”.

⁴ Note that similarity is computed on a trial-by-trial basis, so while inter-item similarity is always maximal between successive examples, it is not necessarily maximized over an entire block (i.e. similarity is maximized locally but not globally).

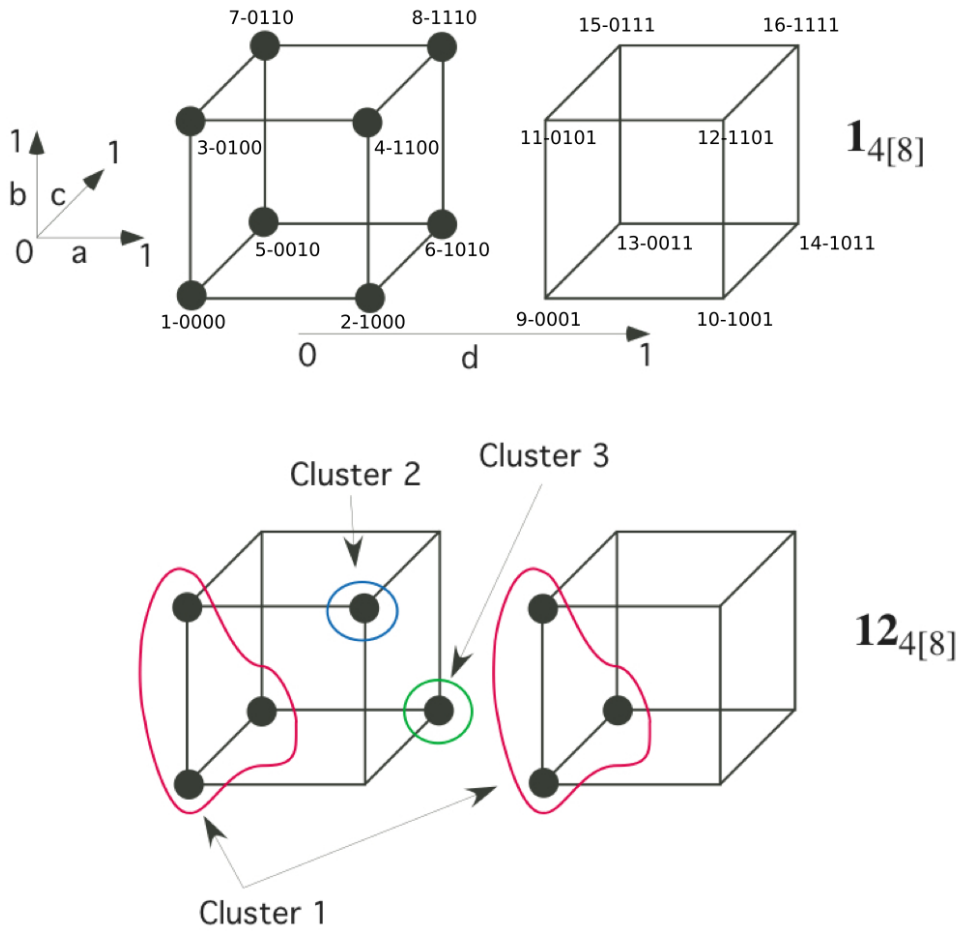


Figure 1. Concepts $1_{4[8]}$ and $12_{4[8]}$ of the $4[8]$ family. (See Feldman, 2000, for further explanation of the concept family taxonomy.) Positive examples are indicated by black circles; negative examples are represented by empty vertices. There is one cluster in concept $1_{4[8]}$ and three clusters in concept $12_{4[8]}$. The stimulus coding order is $abcd$. The code 0000 stands for $a'b'c'd'$, 1111 stands for $abcd$. The number preceding the code is a simpler identification number.

Results

The average inter-item similarity within the similarity-based order, the dissimilarity-based order and the rule-based order was respectively 2.9, 1.5, and 2.3 for concept $1_{4[8]}$ and 2.7, 1.3, and 2.1 for concept $12_{4[8]}$ ⁵. These empirical measures correspond to what would be expected in principle. For example, in concept $1_{4[8]}$, the similarity between two positive stimuli is most of the time equal to 3 in the similarity-based order (once in a while, it can be equal to 2 or 1 at the end of certain blocks, when the path does not allow any other choice), except between the last stimulus of a given block and the first stimulus of the next block (the first stimulus of

a block being drawn randomly). The average inter-item similarity between the last stimulus of the n^{th} block and the first of the $(n + 1)^{\text{th}}$ block is 2.5 (i.e., the average of 4, 3, 3, 3, 2, 2, 2, 1). Therefore the theoretical average inter-item similarity in concept $1_{4[8]}$ is $(7 \times 3 + 1 \times 2.5)/8 \approx 2.9$, agreeing with the empirical measure. These measures will be used for comparison purposes in the following analysis.

We first consider the influence of presentation order on learning, and then provide more detailed analyses of the pro-

⁵ The average inter-item similarity was computed for positive examples only because negative examples were presented in random order in all presentation orders.

Table 1
Encoded study items presented in Fig. 1, and presentation order samples in concept 1_{4[8]}

1 _{4[8]}		12 _{4[8]}		Presentation order samples in 1 _{4[8]}					
#	Cat 1	#	Cat 1	SBO	DBO	RBO			
1	0000	1	0000	1	0000	-	Cat 0	3	0100
2	1000	3	0100	2	1000	8	1110	5	0010
3	0100	4	1100	-	Cat 0	-	Cat 0	7	0110
4	1100	5	0010	4	1100	1	0000	-	Cat 0
5	0010	6	1010	-	Cat 0	-	Cat 0	-	Cat 0
6	1010	9	0001	3	0100	4	1100	8	1110
7	0110	11	0101	7	0110	5	0010	4	1100
8	1110	13	0011	8	1110	-	Cat 0	-	Cat 0
				-	Cat 0	3	0100	1	0000
				-	Cat 0	6	1010	-	Cat 0
				6	1010	-	Cat 0	-	Cat 0
				5	0010	-	Cat 0	2	1000
				-	Cat 0	7	0110	6	1010
				-	Cat 0	-	Cat 0	-	Cat 0
				-	Cat 0	2	1000	-	Cat 0
				-	Cat 0	-	Cat 0	-	Cat 0
#	Cat 0	#	Cat 0						
9	0001	2	1000						
10	1001	7	0110						
11	0101	8	1110						
12	1101	10	1001						
13	0011	12	1101						
14	1011	14	1011						
15	0111	15	0111						
16	1111	16	1111						

Note. In the “Presentation order samples in 1_{4[8]}” column, Cat 0 cells can be replaced by any negative examples of the Cat 0 category, because negative examples were drawn in random order; SBO, Similarity-based order; DBO, Dissimilarity-based order; RBO, Rule-based order; Stimulus # is also indicated in Fig. 1

gression of learning over time, and finally of classification response times.

Influence of presentation order on learning

Among the 96 subjects (32 by presentation order), only 69 subjects could finish the experiment within the time slot allocated to the experiment. The analysis of results only takes into account these 69 subjects (25 for the rule-based procedure, 21 for the dissimilarity order, and 23 for the similarity order). The loss of subjects apparently tracked the difficulties associated with the presentation orders presented below, although the chi-square test of independence between presentation order and loss was not significant, $\chi^2(2) = 1.2, ns$.⁶

Figure 2 shows the number of blocks which were required for subjects to reach the learning criterion in the three conditions (rule-based, similarity-based, and dissimilarity-based) for both concepts (1_{4[8]} and 12_{4[8]}); respectively 5.2, 6.1, and 6.9 blocks for concept 1_{4[8]} and 22.1, 28.1, and 42.6 blocks for concept 12_{4[8]}. The results indicate that there was an effect of the presentation order on learning: the number of blocks required to reach the learning criterion depended on the presentation order that was chosen ($F(2, 66) = 15.3, p < .001, \eta_p^2 = .32$). Learning was fastest in the rule-based order (mean = 13.7 blocks until criterion, s.d. = 10.4), second

fastest in the similarity-based condition (mean = 17.1 blocks, s.d. = 12.6), and slowest in the dissimilarity-based condition (mean = 24.7 blocks, s.d. = 22.5). The superiority of the similarity over dissimilarity order replicates earlier findings. But the main result, that the rule-based ordering is superior to either, is novel.

As can be seen in the Fig. 2, concept 12_{4[8]} was learned much more slowly overall ($F(1, 66) = 320, p < .001, \eta_p^2 = .83$). The effect of presentation order was far larger in magnitude in this complex concept, which is reflected in the interaction between concepts and presentation orders ($F(2, 66) = 15.8, p < .001, \eta_p^2 = .32$). When analyzing simple effects of presentation orders for each concept, the effect of presentation order was only significant in concept 12_{4[8]} ($F(2, 66) = 16.34, p < .001, \eta_p^2 = .33$). The between-subjects *t* tests indicated that only the three paired comparisons between presentation orders for concept 12_{4[8]} were significant. However, a subsequent analysis of learning curves will reveal that presentation order also influences

⁶ The loss of subjects probably stems from our subject scheduling system, which unfortunately did not leave many subjects sufficient time to complete the experiment given the strict 100% criterion. Elio and Anderson in their second study also excluded 14 cases among eighty subjects with a 85% correct criterion.

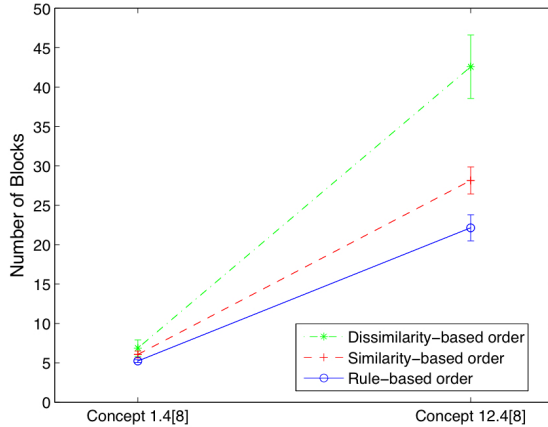


Figure 2. Number of blocks taken to reach the learning criterion of 100% correct classification for two consecutive blocks. Error bars show +/- one s.e.

learning of concept $1_{4[8]}$.

Detailed analysis of the progression of learning

We next turn to the question of how learning progresses over time in the three presentation orders. Fig. 3 (concept $1_{4[8]}$) and Fig. 4 (concept $12_{4[8]}$) show both the percentage and the number of correct responses for each block as a function of block number over the course of the experiment. Note that the absolute number is sometimes more revealing than the percentage; for example 14 correct responses (87.5%) can be immediately understood to mean “all but two in the concept.”

Concept $1_{4[8]}$. As can be seen in Fig. 3, learning of the simpler concept $1_{4[8]}$ was slightly more efficient in the rule-based than in the similarity order. The fit comparison procedure described in the notes of Fig. 3 indicated that the two curves differ significantly in fitted form ($F_{RS}(42, 44) = 13.17, p < .001$). Thus even in concept $1_{4[8]}$, though the difference is relatively subtle, the rule-based order produces significantly more rapid learning than the similarity order. In contrast, the dissimilarity order seems to have induced ineffective learning compared to the similarity-based order ($F_{DS}(42, 44) = 19.25, p < .001$) and the rule-based order ($F_{DR}(42, 44) = 18.46, p < .01$).

Concept $12_{4[8]}$. Using a similar fit comparison procedure, the same ranking of effectiveness of the three presentation orders (rule-based > similarity > dissimilarity) was visible in the learning curves in concept $12_{4[8]}$ (again much larger in magnitude than in the simpler concept; we obtained $F(144, 147) > 40, p < .001$ for the three paired comparisons of the learning curves).

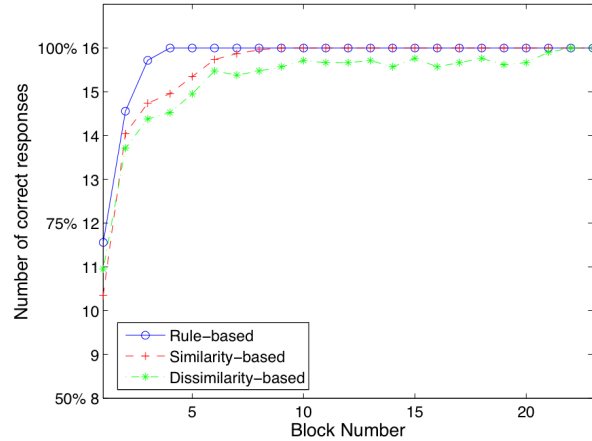


Figure 3. Performance as a function of block number for the three presentation orders, concept $1_{4[8]}$. To compare any pair of learning curves, we fitted the data sets to a common nonlinear model $y = b_0 + b_1/x$ which fit all three datasets well ($R^2 > .89$ in all cases). We tested the null hypothesis of equal slopes between pairs of regression curves by comparing the mean squared error when the two datasets were pooled compared to when they were fit separately. (the technique is similar to verifying that there is no interaction between the covariate and treatments before running an ANCOVA). The results showed that the three learning curves are statistically distinct.

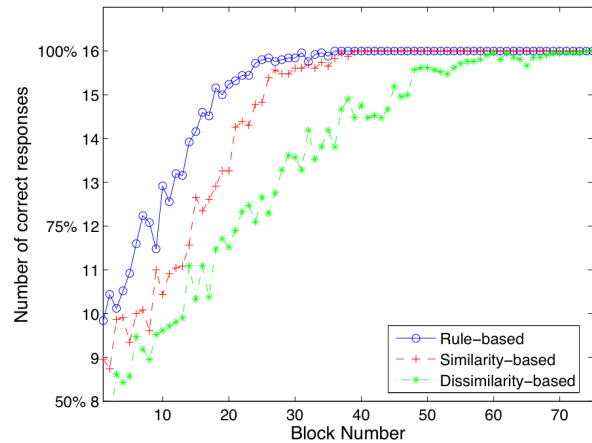


Figure 4. Performance over time (blocks) for the three presentation orders, concept $12_{4[8]}$. The three learning curves are statistically distinct. The fit comparison procedure is described in the notes of Fig 3. This time we fitted the data sets to a quadratic model $y = b_0 + b_1x + b_2x^2$ which fit all three datasets well ($R^2 > .92$ in all cases).

Discussion

Like several previous studies, this study demonstrates that the sequence in which examples are encountered can profoundly influence the success of learning. Our results show that a rule-based order yields learning superior to the similarity order previously found most advantageous (Medin & Bettger, 1994). Though our interest is primarily theoretical, this result has obvious implications for the presentation of material in educational settings (Avrahami, 1987).

Our results demonstrate that the benefit of the rule-based order does not derive entirely from the inter-item similarity it entails. Inter-item similarity was maximal in the similarity-based order, and only intermediate in the rule-based order, but the rule-based order elicited the best performance. We conclude that the nature of the rule-based order (i.e., randomness and clustering) provides an independent learning benefit above and beyond that provided by inter-item similarity. Nevertheless, in order to affirm that the nature of rule-based orders is critical in the performance we observed, subsequent experiments are needed to show more precisely that with an equivalent inter-item similarity, subjects would perform worse in a non rule-based order than in a rule-based order. Note that a reduction in between-block order variability (as a consequence of presenting the six positive items of the largest cluster before the two remaining exceptions on every block) in concept 12_{4[8]} for the rule-based order is very unlikely a source of facilitation for subjects. There are $6! = 720$ possible orders for the six examples of the largest cluster. Subjects could not rely on a reduced number of possible orders to come up with a strategy.

This superiority of the rule-based order over the similarity order might be attributed to the illusory structure that the similarity order might tend to induce in the minds of learners. The similarity condition entails a relatively orderly trajectory through the space that is, in fact, not genuinely informative about the category, and thus might temporarily mislead the learner about the structure to be learned. By contrast the rule-based order, by definition, randomizes that which is not informative (steps within a cluster) while segregating the clusters, yielding superior learning.

These results suggest that human learning does involve a process of rule-based abstraction, consistent with many recent hybrid models (Anderson & Betz, 2001; Erickson & Kruschke, 1998; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Nosofsky, Palmeri, & McKinley, 1994; Rosseel, 2002; Smith & Sloman, 1994) or clustering models such as SUSTAIN (Love et al., 2004), which overtly involve a rule-like component. (SUSTAIN also involves an incremental functioning which can be responsible for over specific solutions when the items are presented in an unfavorable order.) It also might be consistent with a pure exemplar-storage schemes (Estes, 1994; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, Gluck, Palmeri, McKinley, & Gauthier, 1994),

as these are also sensitive to the homogeneity of categories (Hintzman, 1986), though the connection would be less direct. In any case, exemplar models are generally “batch” models (i.e., in batch algorithms, categorization probabilities are not computed trial-by-trial, but for the whole set of examples block after block), meaning that they consider the set of examples as an unordered group, and hence could not model our results without some sort of extension. One exception is ALCOVE (Kruschke, 1992), because it includes trial-by-trial updating. AMBRY (Kruschke, 1996), derived from ALCOVE, might also be adequate for fitting our results, because it allows changes of category-to-response association weights and exemplar-to-category association weights after every trial in order to model rapid shifting in categorization. We feel that the superiority of rule-based presentation orders is a key test for models of categorization learning: a finding that any competitive model needs to be able to account for in principle. The comparative benefits of different order may be helpful in deciding among competing rule-based models, which differ in nature of the rules extracted (Bradmetz & Mathy, 2008; Feldman, 2000; Lafond, Lacouture, & Mineau, 2007; Vigo, 2006 and also Love et al., 2004; Nosofsky, Palmeri, & McKinley, 1994). That is, alternative rule-based orders could be devised to match the abstraction or compression techniques entailed by the various theories. Following our argument, the most effective presentation orders for subjects would be that which accords with the subject’s own internal hypotheses or representations.

Of course, there is no guarantee that any given presentation order induces in subjects anything like the mechanisms involved in its construction. A rule-based order does not necessarily induce the formation of rules, nor does a similarity-based order necessarily induce the computation of similarity; nor for that matter does a random presentation necessarily induce anything like random guessing or rote memorization. The learning mechanisms hypothesized in rule-based models and those hypothesized in exemplar models both capitalize on the structure present in the observations, though in different ways.

The differences among presentation orders depend in an interesting way on the nature of the concepts learned. Most obviously, the rule-based presentation order only provides a substantial benefit if the category is highly structured—that is, when it contains salient subcategories around which the presentation order can be organized. For concepts like 1_{4[8]} that lack internal subdivisions, a rule-based presentation is in effect a random presentation. But a similarity order might induce a sequence of temporary over-specific hypotheses (blind alleys) based on accidentally contiguous examples, which would impede learning.

We noted a particular negative effect of the dissimilarity-based order, in which subsequent positive examples were chosen as distant as possible from each other. Research on

the effect of the relative magnitude information may account for this result (Stewart, Brown, & Chater, 2002). Stewart et al. showed that categorization of a stimulus on trial n is influenced by the stimulus and response on trial $n - 1$, when information about presentation orders is not discarded. Their Memory and Contrast (MAC) model predicts that participants tend to respond with a different category when the difference between two consecutive examples is large.

Conclusion

We believe that the current investigation is an important step in understanding the effect of presentation order, but several new conditions need to be explored to gain additional insight. A first extension would be to separate the training phase (in which the presentation order is manipulated) from the categorization phase. This would allow the positive examples to be presented alone, rather than being interspersed with negative examples, which muddies the desired order.

We also plan to model our results, both by existing incremental models, as well as by extensions of existing models. Exemplar models might be extended to take the temporal dimension into account by including sequential order as a feature in similarity comparisons. The similarity between two stimuli would be influenced by their relative serial position, inducing a neighborhood structure in terms of both features and time. The model could then capitalize on the temporal dimension to assess local distinctiveness, in the same way serial position effects are modeled as discrimination problems in serial or free recall (Brown, Neath, & Chater, 2007). Another solution proposed by Stewart et al. (2002) is to adapt an exemplar model to predict sequence effects by weighting the stimulus on the previous trial more heavily than others in the summed similarity calculations, on the grounds that recent stimuli are more available in memory, or simply that they ought to be weighed more heavily in decisions.

References

- Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, 8(4), 629–647.
- Avrami, J. (1987). Teaching by examples: Implications for the process of category acquisition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 50A, 586–606.
- Bradmetz, J., & Mathy, F. (2008). Response times seen as decomposition times in Boolean concept use. *Psychological Research*, 72, 211–234.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114, 539–576.
- Clapper, J. P., & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 443–460.
- Elio, R., & Anderson, J. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, 12, 20–30.
- 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.
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107–140.
- Estes, W. K. (1994). *Classification and cognition*. New York, NY: Oxford University Press.
- 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.
- 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.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93(4), 411–458.
- 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.
- Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology*, 51, 57–74.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: A network model of category learning. *Psychological Review*, 111, 309–332.
- Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin and Review*, 1, 250–254.
- Medin, D. L., & Schaffer, M. (1978). A context theory of classification learning. *Psychological Review*, 85, 207–238.
- 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.
- Rossee, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, 46(178–210).
- Sakamoto, Y., Jones, M., & Love, B. C. (2008). Putting the psychology back into psychological models: Mechanistic versus rational approaches. *Memory & Cognition*, 36(1057–1065).
- Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, 133, 534–553.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75, 13, whole No. 517.

- Smith, E. E., & Sloman, S. A. (1994). Similarity- vs. rule-based categorization. *Memory & Cognition*, 22(4), 377–386.
- 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.
- Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology*, 50, 501-510.