Order Effect and Time Varying Categories

Lee-Xieng Yang (lxyang@nccu.edu.tw)
Department of Psychology, Researcher Center for Mind, Brain and Learning
National Chengchi University, No.64, Sec.2, ZhiNan Rd., Taipei City
11605, Taiwan (R.O.C)

Hao-Ting Wang (htwangtw@gmail.com)
Department of Psychology, National Chengchi Univserity
No.64, Sec.2, ZhiNan Rd., Taipei City
11605, Taiwan (R.O.C)

Abstract

The main purpose of this study is to examine how people learn time-varying categories as well as whether order effect exists in such learning. To this end, we design three types of category structures, in which the stimuli vary along trials in an ascending, descending, and quadratic trend. Also, tendency to repeat preceding category label as current response is regarded as evidence for order effect. The results show a clear order effect in these experiments. The modeling results reveal that GCM, which is modified to be sensitive to trial order and SDGCM, which relies on the similarity and dissimilarity to the exemplars for categorization, provide a good fit for all experiments. However, the rule-based model used by Navarro, Perfors, and Vong (2013), which changes the boundary trial by trial has difficulty accommodating the learning pattern in quadratic trend.

Keywords: Order Effect; Time-Varying Category; Category Learning

Time-Varying Categories

Many natural categories are characterized by their features varying in magnitude with time. For instance, the leaves of the deciduous plant are green in summer and gradually turn to yellow or red in autumn. This type of categories we call time-varying categories. Recently, Navarro and colleagues used artificial time-varying categories to examine how people learn on them. They found that although the task was not easy, people could not only get a satisfactory learning accuracy, but also predict the forthcoming valid items based on their understanding of the category structure (Navarro & Perfors, 2009, 2012; Navarro, Perfors, & Vong, 2013).

Also, these authors showed that their results can be accommodated by the models, which either give the recently seen items a larger weight in the framework of the exemplar or the prototype model (Navarro & Perfors, 2012) or shift the boundary trial by trial in the framework of a rule model (Navarro et al., 2013). Whichever way it goes, the key to making a model capable of accounting for the learning of time-varying categories is the sensitivity to time, specifically the sensitivity to trial order.

Order effect

The request for the sensitivity to trial order for modeling learning pattern of time-varying categories implies that some sort of order effect should be expected in learning of time-varying categories. Order effect can be broadly defined as any influence on current item brought by preceding item during learning. Herein, order effect is referred to as tendency of repeating category label of preceding item as response for current item.

The simulation work of Stewart, Brown, and Chater (2002) following their MAC (Memory and Comparison) strategy is a good demonstration for order effect. In their simulation experiment, 10 items were divided into two categories, 5 taking lower values and the other 5 taking higher values on the stimulus dimension. Suppose item X_{n-1} is from the low category, when item X_n takes a value even lower than X_{n-1}, it must come from the low category and vice versa when item X_{n-1} is from the high category. That is, the probability to repeat the category label of X_{n-1} is 1.00.

However, when the change on stimulus value from preceding item to current one cannot guarantee what the current category label is, that is when item X_{n-1} is from the low category and $X_n > X_{n-1}$ or when item X_{n-1} is from the high category and $X_n < X_{n-1}$, the probability to repeat the category label of item X_{n-1} depends on the similarity between items X_{n-1} and X_n.

These authors showed that MAC strategy, thought very simple, can get about 80% accuracy on predicting participants' performance, hence highlighting that the exemplar temporarily retained in STM is sufficient to provide reliable information for categorization.

In this study, we would like to extend the studies of Navarro and Perfors (2012) and Navarro et al. (2013) and attempt to reveal the potential order effect in learning of time-varying categories. To this end, we conduct three experiments with different category structures and compare three computational models on fit to the collected data. Before we introduce the experiments, we first introduce the models we would like to test in this study.

GCM

GCM (Generalized Context Model) (Nosofsky, 1986, 1987) is a classic exemplar-based model, positing that the inter-item similarity is the basis of categorization and any item would be classified as the category to which it is more similar. The similarity between each item I and exemplar J, $S_{I,J}$, is the negative exponential function of distance between them on the stimulus dimension [1], which is derived as

$$S_{I,J} = exp^{-c|X_I - X_j|}, \tag{1}$$

where c is the specificity parameter, the larger c is the more unique the items are in the psychological space.

The similarity of item I to Category 1, $S_{I,1}$, is the sum of the similarities to the exemplars of Category 1, $S_{I,1} = \sum_{j\in 1} S_{I,j}$, and so is the similarity to Category 2, $S_{I,2} = \sum_{j\in 2} S_{I,j}$. The probability of Category 1 on item I is computed as

$$p(1|I) = \frac{\beta S_{I,1}}{\beta S_{I,1} + (1-\beta)S_{I,2}}, \tag{2}$$

where β is the bias to Category 1, normally set as .5 when the two categories have equal amount of the exemplars.

Although the original GCM is not sensitive to the trial order, it can be adjusted by decreasing the weighting for the exemplars exponentially from trial $n-1$, trial $n-2$, until trial 1 (see Nosofsky, Kruschke, & McKinley, 1992; Nosofsky & Palmeri, 1997). Thus, in this study, the similarity between item I and exemplar J is weighted by $w_J = exp^{\lambda(j-(n-1))}$, where j is the trial number of exemplar J starting backwards from $n-1$. The category similarity to Category 1 hence becomes $S_{I,1} = \sum w_j S_{I,j\in 1}$ and so does the category similarity to Category 2, $S_{I,2} = \sum w_j S_{I,j\in 2}$. In this study, the specificity c and the decay rate of the weighting for the exemplar λ would be freely estimated while fitting the GCM to the collected data. The bias to category β is fixed as .50, therefore only the similarity to category matters when predicting the category.

SDGCM

The second exemplar-based model we would like to test is SDGCM (Similarity-Dissimilarity Generalized Context Model) (Stewart & Brown, 2005), in which not only the similarity to the exemplars but also the dissimilarity to them is taken into account. Thus, the similarity of item I to Category 1, $\nu_{I,1}$, is the weighted sum of the similarities to the exemplars of Category 1, $S_{I,j\in 1}$, and the dissimilarities to the exemplars of Category 2, $1 - S_{I,i\in 2}$,

$$\nu_{I,1} = \sum_{j\in 1} w_j S_{I,j\in 1} + \sum_{j\in 2} w_j (1 - S_{I,j\in 2}). \tag{3}$$

The parameter w_j adjusts how much the exemplar j contributes to the computation of the category similarity and it follows the same exponential function described in the section of GCM. The probability of Category 1 is computed as

$$p(1|I) = \frac{(\beta\nu_{I,1})^{\gamma}}{(\beta\nu_{I,1})^{\gamma} + ((1-\beta)\nu_{I,1})^{\gamma}}, \tag{4}$$

where β is the bias to Category 1 and γ is a parameter that varies the degree of determinism in responding. For the sake of simplification, β is set as .5 and γ is set as 1 when modelings the data of this study. Thus, same as the GCM, there are two freely estimated parameters (i.e., c and λ) for SDGCM. The MAC strategy is a special case of SDGCM (see Stewart & Brown, 2005) and this is why we choose SDGCM for modeling in this study.

CBS

Navarro et al. (2013) showed that a rule-based model with the rule shifting trial by trial can predict the learning pattern on the categories. The category boundary linearly rises along the trial order. Herein, we call their model CBS (Category Boundary Shifting) model. The idea of CBS model is very simple. People estimate the location of the category boundary for item n, $\widehat{\mu}_n$, on trial $n-1$, by shifting the boundary for item $n-1$, $\widehat{\mu}_{n-1}$, by some proportion w, $0 \leq w \leq 1$, towards item $n-1$. The boundary at the current moment is always the cumulation of all previously seen items, yielding

$$\widehat{\mu}_n = w \sum_{j=1}^{n} (1-w)^{j-1} X_{n-j} + b. \tag{5}$$

It is obvious in Equation 5 that the recent items would be weighted more heavily. Also, when w is large, only a few preceding items would be used. On the other hand, when w is small, the items seen earlier would also contribute to the estimation of current boundary. The parameter b is used to avoid the model prediction lagging too much behind the data. Once the boundary is estimated, the difference between the current item and the boundary Δ_n is transferred by a sigmoid function to the probability of a category, say Category 1[2], which is given by

$$p(1) = \frac{1}{1 + exp^{-\lambda\Delta_n}}. \tag{6}$$

The parameter λ determines the steepness of the sigmoid function, the larger λ the more deterministic the classification of the items would be. When fitting CBS to the data in this study, three parameters, w, λ, and b, are freely estimated.

In Equation 6, if the item is larger than the boundary, then it is impossible to get the probability of Category 1 less than .50. If the item is smaller than the boundary, it is impossible to get the probability of Category 1 more than .50. Presumably, the CBS model is only suitable for the case where the category relationship stays constant. However, the other two exemplar-based models have no such constraint.

Overview of Experiments

There are two goals in this study. First, we would like to examine order effect in the learning of time-varying categories. Three experiments are conducted with the lines of different lengths as the stimuli. The stimulus lines for learning are manipulated as always increasing in length (the ascending cat-

[1]As all category structures are one dimensional in our experiments, herein the GCM is simplified and has no selective attention weight. The distance is measured in the city-block metric when doing the modeling of the exemplar-based models in this study.

[2]Category 1 is defined as the high category on the stimulus dimension.

egory structure), decreasing in length (the descending category structure), and decreasing first and then increasing (the quadratic category structure). These category structures can be seen in Figure 1, which are deployed in separate experiments.

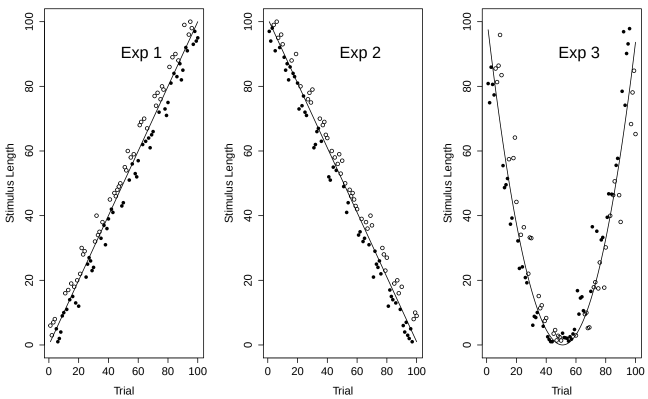

Figure 1: The category structures used in the three experiments. The abscissa shows the trial number and the ordinate shows the stimulus length. From left to right: ascending, descending, and quadratic.

Second, we would like to compare the fit of the three reviewed models to the learning data. Specifically, we place the emphasis on how these models can account for order effect discussed in the previous section. To this end, we divide the trials in each experiment to eight types: according to the category label of the preceding item (Category 1 vs. Category 2), the direction of length change from the previous item (Up vs. Down), and in which learning session the current trial is (First half vs. Second half). Although only Experiment 3 is expected to have different patterns of order effect in the first and the second half session of learning, for a fair comparison, we keep this division for the first two experiments. The goal of parameter optimization is to increase the likelihood of each model to predict the observed probability of Category 1 on these types of trials.

For fairness, the model using more parameters is penalized. The goodness of fit is measured as $AIC = -2LogL + 2N$, N = the number of freely estimated parameters (Akaike, 1974). The log likelihood is $LogL = \sum_i log(\sum_k f_{ik})! - \sum_i \sum_k (log f_{ik})! + \sum_i \sum_k f_{ik} log(p_i k)$, where k is the number of categories (i.e., $k = 2$) and i is the number of the types of trials (i.e., $i = 8$).

Experiment 1

The category structure used in this experiment is shown in the left panel of Figure 1. The white circles denote Category 1, which is always larger than Category 2 denoted by black circles. The categorization rule "Respond Category 1, if $X_n > n$ and Category 2 otherwise", where n is the trial number from 1 to 100, is followed.

This category structure is designed the same way as used by Navarro et al. (2013). However, our stimuli are displayed in a manner of block randomzation. That is the total 100 lines from short to long are separated to 10 blocks. Within each block, the order of the stimuli is random, but the block sequence is sorted in an ascending way. As a result, this design would make our task harder than theirs to learn.

Method

Participants In total, 25 undergraduate students are recruited to participate in this experiment. The task can be finished in less than 20 minutes on average. The participant, after finishing the task, would be reimbursed with NTD$ 60 ($\simeq$ US$ 2) for their time and traffic expense.

Materials and Procedure The stimuli are the horizontal lines of different lengths. The length of the line is transferred to be pixels from the stimulus value (from 1 to 100 units) on the dimension by $\frac{screen-width(pixels)}{screen-width(cm)} \times (0.3 + 0.2value)$. On each trial, a horizontal line is presented on the center of screen. When the participant make a response by pressing the key s or ' for Category 1 and Category 2 respectively, a corrective feedback is presented (correct or wrong). After 1 sec, the next trial begins.

Results

See the solid lines in Figure 2 for the learning accuracy. The black circles represent the probability correct made by the participants. Apparently, Category 1 (mean accuracy = .74) is learned better than Category 2 (mean accuracy = .49). This result is consistent with the finding of Navarro et al. (2013) (.70 vs. .60 for Category 1 and Category 2). A Category (2) $\times$ Block (10) within-subject ANOVA confirms this inspection [$F(1,24) = 59.98$, $MSe = 0.13$, $p < .01$] and shows that the learning performance is marginally significant throughout the blocks [$F(9,216) = 1.851$, $MSe = 0.05$, $p = .06$] and that the learning pattern is not different for different categories [$F(9,216) = 1.39$, $MSe = 0.04$, $p = .19$]. It suggests that Category 1 is learned better than Category 2 throughout the whole learning session.

In order to examine order effect, a Category (2) $\times$ Direction-Change (2) $\times$ Learning-Session (2) within-subject ANOVA is conducted for the probability of Category 1 on 99 trials[3]. The results show a significant main effect of Category [$F(1,24) = 39.86$, $MSe = 0.11$, $p < .01$], a significant main effect of Direction-Change [$F(1,24) = 23.83$, $MSe = 0.07$, $p < .01$], and no significant main effect of Learning-Session [$F(1,24) = 3.79$, $MSe = 0.03$, $p = .06$]. The tendency to make a response of Category 1 depends on the category label of the preceding item as well as the direction on the length changing from the preceding item to the current one [$F(1,24) = 6.06$, $MSe = 0.02$, $p < .05$]. In different learning sessions, the tendency of making a Category 1 response differs when the length change differs in di-

[3] The first trial is omitted, as order effect begins from the second trial.

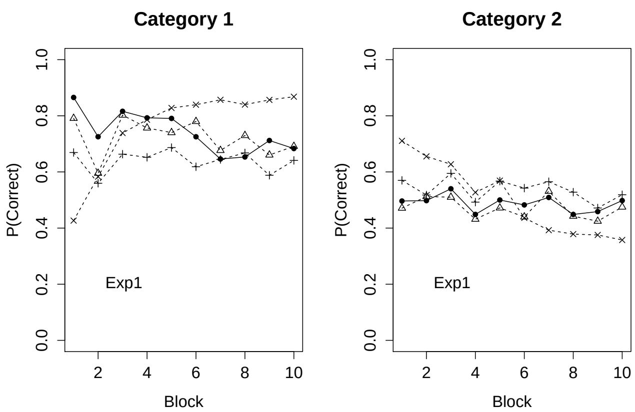

Figure 2: The observed and predicted P(Correct) in Experiment 1.

rection [$F(1,24) = 11.9$, $MSe = 0.02$, $p < .01$]. However, there is no interaction effect between Category and Learning-Session, $F(1,24) < 1$, nor the three-way interaction effect [$F(1,24) = 1.25$, $MSe = 0.02$, $p = .028$].

The superiority of learning Category 1 over Category 2 may be viewed as evidence for a response bias toward Category 1. However, the significant interaction effect between Category and Direction-Change turns down this hypothesis. There would be no such interaction effect, had the bias existed. Following the view of order effect, the asymmetry on learning accuracy for the two categories can be explained. As the lines are getting longer, there would be a greater chance to see the case when a line is longer than its preceding line from Category 1 than when a line is shorter than its preceding line from Category 2. Thus, repeating the preceding answer for the current item would make the trial more correct for Category 1 than Category 2. Therefore, the better learning in Category 1 instead suggests the existence of order effect in this experiment.

Experiment 2

In this experiment, we put emphasis on examining how people learn the descending structure. See the middle panel in Figure 1. The categorization rule is same as the one used in Experiment 1: "Respond Category 1, if $X_n > n$ and Category 2 otherwise". As in Experiment 1, order effect is expected. Also, since the lines are getting shorter and shorter, according to order effect, Category 2 is expected to be learned better than Category 1 instead in this experiment.

Method

Participants In total, 25 undergraduate students are recruited to participate in this experiment. The whole task can be finished in less than 20 minutes on average. After testing, the participants would be reimbursed with NTD$ 60 ($\simeq$ US$ 2) for their time and traffic expense.

Materials and Procedure The stimuli and the testing procedure are the same as used in Experiment 1.

Results

The overall accuracy on Category 1 is worse than that on Category 2 (.53 vs. .75). This inspection is consistent with our expectation and confirmed by a Category (2) $\times$ Block (10) within-subject ANOVA. The results reveal that Category 1 is learned significantly worse than Category 2 [$F(1,24) = 61.95$, $MSe = 0.10$, $p < .01$], the learning block has a significant main effect [$F(9,216) = 2.12$, $MSe = 0.06$, $p < .05$], and Category 2 is learned better no matter which block it is in [$F(9,216) = 1.60$, $MSe = 0.04$, $p = .12$]. The learning pattern across blocks is shown by the solid line in Figure 3. The black circles connected by solid lines represent the observed probability of correct.

In order to test order effect, a Category(2) $\times$ Direction-Change (2) $\times$ Learning-Session (2) within-subject ANOVA is conducted for the probability of Category 1 on 99 trials. The results show that the participants make more Category 1 response when following an item from Category 1 (mean P(1) = .55) than when following an item from Category 2 (mean P(1) = .26) with $F(1,24) = 78.58$, $MSe = 0.06$, $p < .01$. Also, the participants show a different tendency to make a response of Category 1 for a longer or shorter current line [$F(1,24) = 43.40$, $MSe = 0.09$, $p < .01$]. However, there is no difference on the probability of Category 1 made for the trials in the first half or the second half learning session [$F(1,24) = 2.64$, $MSe = 0.01$, $p = .12$].

As Experiment 1, the interaction effect between Category and Direction-Change is significant [$F(1,24) = 5.43$, $MSe = 0.02$, $p < .05$] and so is the interaction effect between Direction-Change and Learning-Session [$F(1,24) = 9.92$, $MSe = 0.02$, $p < .01$]. The tendency of making a Category 1 response is not different in different learning session. The three-way interaction effect is not significant [$F(1,24) = 1.32$, $MSe = 0.02$, $p = .26$]. Again, it is revealed that the tendency of making a Category 1 response depends on the preceding item's category label as well as the change on the line length, namely order effect.

Experiment 3

Different to the previous two experiments, the length of the stimulus line in this experiment is designed to vary in a quadratic trend throughout the learning trials. The category structure is shown in the right panel in Figure 1. For the trials before trial 51, the line length is getting shorter and the item of Category 1 is longer than that of Category 2. However, from trial 51 to trial 100, the line length is getting longer and the item of Category 1 is shorter than that of Category 2. With this category structure, we would like to examine whether people can learn the categories whose members vary in time in a quadratic trend as well as whether order effect occurs.

According to the previous two experiments, the descending trend would favor the category with shorter lengths (i.e.,

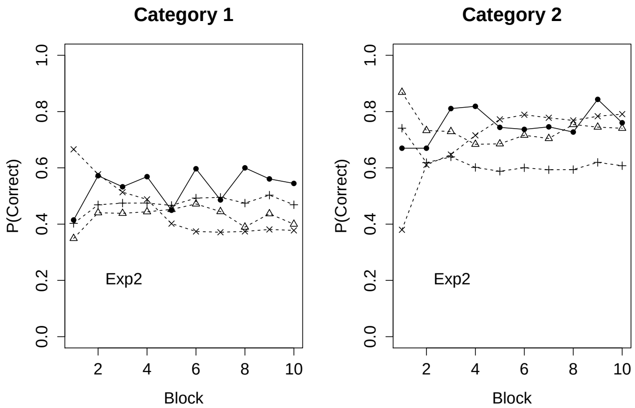

Figure 3: The observed and predicted P(Correct) in Experiment 2.

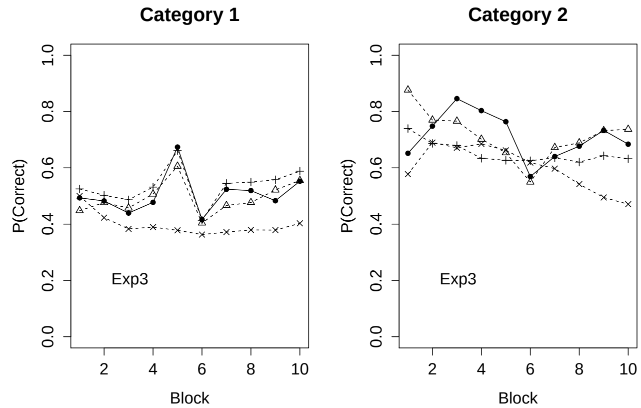

Figure 4: The observed and predicted P(Correct) in Experiment 3.

Category 2 in the first half of the learning session) and the ascending trend would favor the one with longer lengths (i.e., Category 2 in the second half of the learning session). Thus, Category 2 is expected to be learned better throughout all the trials in this experiment.

Method

Participants In total, 25 undergraduate students are recruited to participate in this experiment. The task would be finished in less than 20 minutes on average. After testing, they would be reimbursed with NTD$ 60 ($\simeq$ US$ 2) for their time and traffic expense.

Materials and Procedure The stimuli and the testing procedure are the same as used in Experiment 1.

Results

A visual inspection on Figure 4 shows that, on average, Category 1 is learned worse than Category 2 (.50 vs. .71). The analysis of variance on the probability of correct response confirms this observation [$F(1,24) = 71.01$, $MSe = 0.07$, $p < .01$]. Also, the probability of correct response is significantly different across the learning blocks [$F(9,216) = 2.94$, $MSe = 0.06$, $p < .01$]. The interaction between Category and Block is significant [$F(9,216) = 2.058$,$MSe = 0.06$,$p < .05$]. There is a sudden drop in the learning curve at the sixth block for both categories. This should be related to the chaos brought by the flip of category relationships.

To test order effect, a Category (2) $\times$ Direction-Change (2) $\times$ Learning-Session (2) within-subject ANOVA is conducted. The results reveal that the probability of Category 1 made for the current trial differs when the preceding category label differs [$F(1,24) = 133.70$, $MSe = 0.05$, $p < .01$]. Whether the current line is longer or shorter than the preceding one has no effect on the tendency of making a category 1 response [$F(1,24) = 3.34$, $MSe = 0.04$, $p = .08$]. The probability of Category 1 does not differ in different learning sessions [$F(1,24) = 3.68$, $MSe = 0.02$, $p = .07$]. The two way interaction effect is not significant between Category and Direction-Change [$F(1,24) = 1.61$, $MSe = 0.02$, $p = .22$] and nor is that between Category and Learning-Session [$F(1,24) < 1$]. However, the interaction effect between Direction-Change and Learning-Session is significant [$F(1,24) = 33.65$, $MSe = 0.06$, $p < .01$]. The three way interaction effect is significant [$F(1,24) = 4.62$, $MSe = 0.03$, $p < .05$].

See the black bars in Figure 5 for the probability of Category 1 on the eight types of trials. In the first half of learning session, the participants strongly predict the current item as Category 1 and Category 2 respectively on the U1 and D2 trials. They however are less certain when the direction of length change does not guarantee the category label (i.e., U2 and D1 trials). This result matches the descending part of the category structure. In the second half of learning session, Category 1 becomes shorter than Category 2. Thus, the D1 and U2 trials instead provide reliable information about the category label. This is why the three way interaction effect is significant.

Modeling and Discussion

The three models discussed in the previous sections are fitted to the observed p(Category 1) on the eight types of trials for each participant in each experiment. We adopt *AIC* as the measure of goodness of fit. The smaller the better. The results show that GCM performs the best ($AIC = 21.44$ for Exp1, $AIC = 24.8$ for Exp 2, and $AIC = 32.49$ for Exp3), SDGCM performs the second best ($AIC = 24.81$ for Exp1, $AIC = 28.02$ for Exp2, and $AIC = 37.04$ for Exp3), and CBS performs relatively worse than the other two ($AIC = 28.78$ for Exp1, $AIC = 27.90$ for Exp2, and $AIC = 49.45$ for Exp3).

The best fitting λ of GCM across all experiments is 0.37, suggesting that the ratio of similarity information for classifying a current item is down to less than 50% on trial n-3. The best fitting λ of SDGCM across all experiments is 2.72, suggesting only the right preceding item is included in

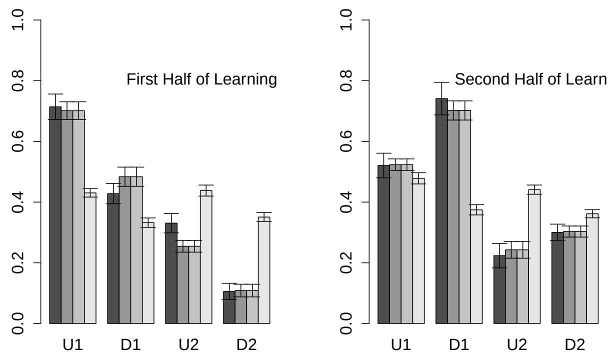

Figure 5: The observed and predicted p(Category 1) for eight types of trials in Experiment 3. U1: item $n >$ item $n-1$ which belongs to Category 1, D1: item $n <$ item $n-1$ which belongs to Category 1, U2: item $n >$ item $n-1$ which belongs to Category 2, and D2: item $n <$ item $n-1$ which belongs to Category 2. The bars from the darkest to the lightest correspond to the observed p(Category 1), the GCM prediction, the SDGCM prediction, and the CBS prediction.

classifying the current one. It is implied that only the temporarily retained item(s) would be needed for learning time-varying categories. The estimated specificity is larger for GCM ($c = 0.61$) than SDGCM ($c = 0.13$). For CBS, the mean best fitting $w = .43$ suggests a medium degree of reliance on the preceding item to determine the current boundary. The decision-making process of CBS is very deterministic ($\lambda = 17.12$). The boundary constant b is estimated as 9.76, 4409.11, and 11645.47 respectively for Experiment 1, Experiment 2, and Experiment 3. The extremely large b prevents the boundary falling down too quickly, which consequently increases responses for Category 2. This benefits the modeling of learning pattern of the descending structure, but not the quadratic structure.

As shown in Figure 5, the participants are aware of the inversion of category relationship (Category 1 $>$ Category 2 $\rightarrow$ Category 2 $>$ Category 1). The U1 and D2 trials are strongly predicted as Category 1 and Category 2 in the first half of learning session and so are the D1 and U2 trials in the second half of learning session. CBS cannot capture this change, because in CBS, Category 1 $>$ Category 2 at all times. In addition, the model predictions are converted to p(Correct)[4], plotted by the dashed lines in Figures 2, 3, and 4. The predicted p(Correct) of GCM, SDGCM, and CBS are denoted respectively by the symbols of triangle, $+$, and $\times$. Although these models are fitted to the data of p(Category 1), the patterns of converted p(Correct) also support that GCM and SDGCM outperform CBS.

[4]When converting to accuracy, $1-$p(Category 1) would be the accuracy for Category 2.

Conclusion

A number of findings in this study are worth noting. We show that people can learn time-varying categories, even when the category structure varies in a quadratic trend. When learning time-varying categories, response for a current item is influenced by preceding category label and the difference between current item and preceding one, namely order effect. This order effect can be accounted for by the exemplar-based models, GCM and SDGCM, with gradient descent weightings for exemplars. Although the learning pattern on time-varying categories can have a rule-based account, a nonlinear category structure would challenge it. In contrast, the exemplar-based account is relatively general to accommodate the learning patterns with different types of category structures.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions in Automatic Control*, *19*, 716-723.

Navarro, D. J., & Perfors, A. (2009). Learning time-varying categories. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of cognitive science society* (p. 414-424). Austin, TX: Cognitive Science Society.

Navarro, D. J., & Perfors, A. (2012). Anticipating changes: Adaption and extrapolation in category learning. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Building bridges across cognitive sciences around the world: Proceedings of the 34th annual conference of the cognitive science society* (p. 809-814). Austin, TX: Cognitive Science Society.

Navarro, D. J., Perfors, A., & Vong, W. K. (2013). Learning time-varying categories. *Memory and Cognition*, *41*, 917-927.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and classification of integral stimuli. *Journal of Experimental Psychological: Learning, Memory and Cognition*, *13*, 87-108.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266-300.

Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, *49*, 403-409.

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