Title
Exemplar Account for Category Variability Effect: Single Category based Categorization

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Yang, Lee-Xieng
Huang, Tai-Lun

Publication Date
2021

Peer reviewed

Exemplar Account for Category Variability Effect: Single Category based Categorization

Lee-Xieng Yang (lxyang@gmail.com)
Department of Psychology, National Chengchi University
No. 64, Sec. 2, ZhiNan Rd., Taipei, Taiwan

Tai-Lun Huang (tailun840724@gmail.com)
Department of Psychology, National Chengchi University
No. 64, Sec. 2, ZhiNan Rd., Taipei, Taiwan

Abstract

The category variability effect is referred to as that the middle item between two categories is more similar to the low-variability category but tends to be classified as the high-variability category, which challenges the exemplar model. We however hypothesized that this effect can result from the use of the single-category strategy in a binary categorization task, specifically when only the low-variability category is referenced for categorization. One experiment was conducted with a recognition task inserted in the categorization task to selectively deepen the processing for the exemplars of the high-variability category, low-variability category, or both categories. The results showed that the strongest category variability effect occurred when the low-variability category was emphasized in the recognition task. The exemplar model SD-GCM provided a good account for the category variability effect, with a large weight for the low-variability category and a small weight for the high-variability category, hence verifying our hypothesis.

Keywords: Category Variability Effect; Similarity and Dissimilarity; Single-Category Strategy

Background

The category variability effect (CVE) was first reported in the seminal study of Rips (1989), in which the participants judged the middle item (e.g., a circular object of 3 inches in diameter) in between two categories in different variabilities (e.g., COIN and PIZZA) more similar to the category with a smaller variability (e.g., COIN), but classified it as the other with a larger variability (e.g., PIZZA). This effect has been regarded as the evidence against the exemplar model, such as the GCM (Nosofsky, 1986), in which an item is always classified as the category with a higher similarity. Past studies have tended to explain this phenomenon in terms of the category variability. Cohen, Nosofsky, and Zaki (2001) found that this phenomenon can be replicated only when the difference on variability between two categories is extremely large (e.g., the low-variability category even had only one exemplar).

Similarly, Stewart and Chater (2002) showed that CVE can occur, when the within-category variability of the exemplars are highlighted (i.e., showing all exemplars to the participants at the same time). Additionally, Perlman, Hahn, Edwards, and Pothos (2012) found when the difference on variability between two categories is less salient, the middle item tends to be classified as the low-variability category, consistent with the exemplar account, whereas it is more likely to be classified as the high-variability category, when the difference is large. Hsu and Griffiths (2010) demonstrated via empirical experiment and Bayesian modeling that CVE can occur in the observational-learning paradigm, in which the prediction for the middle items between the two category distributions depends on the difference between the likelihoods to generate the same feature given different categories, namely p(feature|category), but not in the feedback-learning paradigm, in which the categorical prediction depends on the estimation of the probability of a category given the feature, namely p(category|feature). The critical middle item in between the two categories, although more similar to the low-variability category, is actually more likely to be generated by the high-variability category, hence inducing CVE. This is also the same reason for the rule-based model GRT (e.g., Ashby & Gott, 1988) to account for this effect, in which the category boundary is assumed to be located on where the two category distributions have the same likelihood to generate a positive instance.

Researchers have tried to make the exemplar model able to predict CVE but have not gained satisfactory results. Nosofsky and Johansen (2000) showed that the GCM can predict CVE, only when the sensitivity parameter c is allowed to be tuned independently for each category. However, the sensitivity parameter represents a global scaling of the psychological space for the specificity of the items in it. When c is large, all items become more dissimilar to each other and vice versa. Thus, it is harmful to the self consistency of the GCM, if different categories in the same psychological space are allowed to have different sensitivities. On the other hand, Sakamoto, Jones, and Love (2008) modified the prototype model to represent a category as a Gaussian distribution with the mean (i.e., prototype) and the standard devision of the category being updated by error-driven learning for optimizing the modeling performance. Of course, this modification renders the exemplar model the capability to predict CVE. However, it also contradicts the basic assumption of the exemplar model that categorization is achieved based on the similarity to the category exemplars (or means). Thus, there so far has not a satisfactory exemplar account for CVE.

Motive of This Study

Despite the models (e.g., the rational model and the GRT) with the assumption that a category can be represented as a probability density function can account for CVE, there are still reasons for us to continuously seek for an exemplar account for this effect. First, in all these models, the probability density function is always assumed to be Gaussian. However, in many studies which reported CVE, the exemplars of a category actually followed a uniform distribution. Although this seems to be of no harm practically, theoretical concerns might be required. On the contrary, the exemplar model with exemplars as category representations can be viewed as a nonparametric model (see Ashby & Alfonso-Reese, 1995), which has no need to assume the type of the pdf and estimate the corresponding parameters.

Second, the failure of the GCM to predict CVE might result from assuming that all exemplars of the two categories are referenced in a binary categorization task. In the GCM, the probability to classify the middle item as the low-variability category $p(L) = S_L/(S_L + S_H)$ is the percentage of the summed similarity of the middle item to the low-variability category S_L in the summed similarity of it to the low- and high-variability categories $S_L + S_H$. As long as $S_L > S_H$, even though they both are quite small, the GCM always tends to classify the middle item as the low-variability category. However, the binary categorization can be done by applying the strategy for learning a single category in an A-or-not-A categorization task. That is, one can treat the exemplars of the low-variability category only as the reference points and classify any item as the low-variability category or not (i.e., the high-variability category), based on the summed similarity to the low-variability category. In this circumstance, the probability to classify the middle item as the low-variability category becomes $p(L) = S_L/(S_L + k)$, where k is the response-criterion parameter (see for equation Zaki & Nosofsky, 2007). When $k > S_L$, $p(L) < 0.5$, hence producing CVE. Of course, it is also possible that the high-variability category is referenced for categorization. That is, $p(H) = S_H/(S_H + k)$. When $k < S_H$, CVE occurs. Therefore, it was hypothesized in this study that the occurrence of CVE might result from the use of this single-category based strategy for categorization.

Categorization with Partial Category Representation

Past studies have provided evidence for that not all exemplars contribute equally to categorization. For instance, in order to enable the exemplar model to predict the sequence effect in category learning, the similarity to the recent exemplars was weighted far more than those less recent ones when making the categorization decisions (see Nosofsky & Palmeri, 1997; Stewart & Brown, 2005). Further, De Schryver, Vandist, and Rosseel (2009) showed that it is not always better to use all of the exemplars than part of them for accommodating human categorization performance. Their Rex LI model, which chooses the best among all combinations of the exemplars as the basis for categorization while modeling the observed data, outperformed the GCM on fit to the data of Nosofsky, Clark, and Shin (1989). Similarly, Vanpaemel and Storms (2008) demonstrated via computer simulation that the suitable representation of a category is neither the exemplars nor the prototype, but the partial abstraction of the category, namely the sub-clusters' means in the category.

Recently, Austerweil, Liew, Conaway, and Kurtz (2019) even showed with empirical experiment and their exemplar model PACKER that when only observing the exemplars of one category, the participants could predict the probability for an item to be classified as the same category or the contrasting category, based on the similarity and dissimilarity to these exemplars. Therefore, it is possible that people can adopt the single-category based strategy for categorization. However, in order to make this study comparable with the past studies, it is not suitable to directly instruct participants to take this strategy for categorization. We will introduce what we did to implicitly induce the use of this strategy in the section of Experiment.

SD-GCM

According to the previous discussion, it is not impossible for the exemplar model to predict CVE, as long as the exemplars of only one category are allowed to be referenced for categorization and the probability to classify an item as the target category is determined by the strength of positive evidence for the target category relative to the strength of all pieces of (positive and negative) evidence accordingly. In fact, the exemplar model the SD-GCM (Similarity and Dissimilarity GCM; Stewart & Brown, 2005) can fulfill all these needs.

The SD-GCM is basically the same as the GCM, except that not only similarity but also dissimilarity is considered for categorization. That is, the valence of the evidence for classifying an item as the target category is computed as the sum of the similarity of it to the target category and the dissimilarity of it to the contrasting category. Also, in the SD-GCM, every exemplar's contribution to classifying an item is weighted by a free parameter representing the unique importance of it for categorization. Thus, for item i, the valence of the evidence for Category A ν_A is computed as

$$\nu_A = \sum_{j \in A} w_j s_{ij} + \sum_{j \in B} w_j ds_{ij}, \tag{1}$$

where s_{ij} and sd_{ij} are respectively the similarity and dissimilarity of item i to exemplar j. The parameter w_j controls to what degree the exemplar j should contribute in categorization, $0 \leq w_j \leq 1$. The similarity s_{ij} between item i and exemplar j is transferred from the distance between them d_{ij} in the same way as the GCM, $s_{ij} = e^{-cd_{ij}}$, where c is the sensitivity parameter and $d_{ij} = \sum_m \alpha_m |x_{im} - x_{jm}|$ when city-block distance is used, with α_m as the attention weight for dimension m. The dissimilarity ds_{ij} is computed as $1 - s_{ij}$. Similarly, the valence of the evidence for Category B is computed as

$$\nu_B = \sum_{j \in A} w_j ds_{ij} + \sum_{j \in B} w_j s_{ij}. \tag{2}$$

The probability to classify an item as Category A is proportional to how much evidence supports Category A in the summed valence of the evidence for the two categories as

$$p(A) = \frac{(\beta_A \nu_A)^\gamma}{(\beta_A \nu_A)^\gamma + (\beta_B \nu_B)^\gamma}, \tag{3}$$

where β is the bias for a category and $\beta_A = \beta_B$ when the two categories have equal numbers of exemplars. The parameter γ shows how deterministic the categorization decision is, the larger the more deterministic.

When all exemplars of one category are weighted the same, the SD-GCM presumably can implement the single-category based strategy. For example, the low-variability category based categorization can be implemented, when the weightings for the exemplars of the low- and high-variability categories are respectively set as extremely large and small, such as $w_L = 1$ and $w_H = 0$. In this case, for classifying item i, the valence of the evidence for the low-variability category is the summed similarity to it $\nu_L = \sum_{j\in L} s_{ij}$. In contrast, the valence of the evidence for the high-variability category becomes the summed dissimilarity to the low-variability category $\nu_H = \sum_{j\in L} ds_{ij}$. Now Equation 3 becomes

$$p(L) = \frac{(\nu_L)^\gamma}{(\nu_L)^\gamma + (\nu_H)^\gamma} = \frac{(\sum_{j\in L} s_{ij})^\gamma}{(\sum_{j\in L} s_{ij})^\gamma + (\sum_{j\in L} ds_{ij})^\gamma}. \tag{4}$$

Comparing to the aforementioned equation in the GCM devised for single-category based categorization $p(L) = S_L/(S_L + k)$, the summed dissimilarity to the low-variability category is functionally equivalent to the response criterion. Similarly, for the high-variability category focused categorization, the decision criterion is the dissimilarity of the to-be-classified item to the high-variability category.

With the dissimilarity to the target category as the decision criterion, it can be expected that CVE will occur more easily when the low-variability category is focused rather than the high-variability category. Given that the middle item is often designed to be distant to either category (i.e., dissimilarity $DS_L >$ similarity S_L), it should be classified as the high-variability category (i.e., CVE), when the low-variability category is focused. Following the same logic, when the high-variability category is focused, the middle item should be classified as the low-variability category. This is not CVE.

However, the possibility should not be precluded that a small sensitivity parameter c might make the similarity larger than the dissimilarity of the middle item to the high-variability category, when the high-variability category is focused. In this case, CVE can also occur. Thus, it is not sure if CVE will occur when the high-variability category is focused. Nonetheless, comparing with the attempt of Nosofsky and Johansen (2000) to estimate c independently for each category, the single-category based strategy is a better solution for the exemplar model to account for CVE, without the harm to the self consistency.

Experiment

This experiment was aimed to gain empirical evidence for our single-category strategy hypothesis. All participants were asked to learn the category structure as shown in Figure 1 and their categorization decision for the middle item (in the midway between the boundary items of the two categories) would be treated as the index for CVE. In addition to the categorization task, inspired by the finding that recognition and categorization might share the same memory system (Curtis & Jamieson, 2018; Jamieson & Mewhort, 2009; Nosofsky, Little, & James, 2012; Nosofsky & Zaki, 1998), a recognition task was inserted in the categorization task, after the training phase and before the transfer phase, with the intention to activate more the exemplars of specific categories. In the Low condition, the participants were given only the exemplars of the low-variability category (without category label) to review and then were asked to judge which among all transfer items was the previously-reviewed (or old) item.

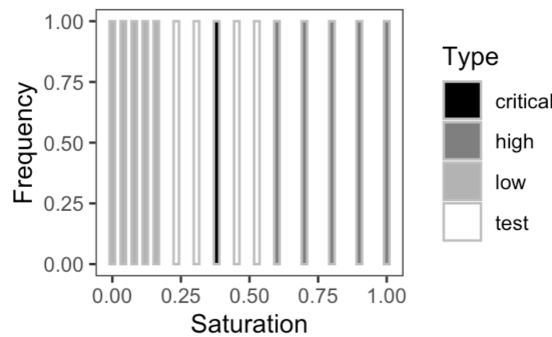

Figure 1: Category structure. The training items of the low- and high-variability categories are denoted respectively by 5 light-grey bars and 5 dark-grey bars. The transfer items consist of the 5 exemplars of each category and the 5 novel items in between with the black bar denoting the critical middle item.

Similarly, in the High condition, the reviewed items were only the exemplars of the high-variability category. In the Both condition, all exemplars were the reviewed items. Presumably, when the exemplars were reviewed in the recognition task, they should be more active than those not reviewed. As the recognition phase was followed by the transfer phase in the categorization task, these reviewed exemplars would be more likely referenced for categorization. Thus, which category should be more relied on for categorization could be manipulated. According to our hypothesis, if this manipulation was successful, it was expected to find a stronger CVE in the Low condition than the other two conditions.

Method

Participants and Apparatus

One hundred forty-three undergraduate students were recruited from National Chengchi University to participate in this experiment. These participants were randomly assigned to the Both condition ($n = 49$), the High condition ($n = 47$), and the Low condition ($n = 47$). The experiment was con-

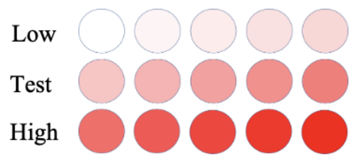

Figure 2: Stimuli for the even-numbered participants. The large saturation values correspond to the high-variability category (denoted by High) and vice versa. Test for the novel items in the transfer phase.

ducted on an IBM compatible PC in a quiet booth for the participants individually. The whole experiment procedure (including stimulus displaying and response recording) was under the control of the script composed with PsychoPy (Peirce, 2007). On average, each participant could finish the test in half an hour. After testing, every participant was reimbursed with NTD$ 120 ($\simeq$ US$ 4) for his/her time and effort.

Stimuli and Procedure

The stimuli were color circles varying on saturation with hue and brightness fixed, which are shown in Figure 2. For counterbalancing the correspondence between saturation values and the two categories, the even-numbered participants were given the stimuli with high saturation for the high-variability category and the odd-numbered participants the stimuli with high saturation for the low-variability category. The saturation values for the even-numbered stimulus set were created by 1 - the values of the odd-numbered stimulus set.

An independent experiment was conducted to examine the correlation between psychological values (i.e., MDS coordinate values in one-dimensional space) and physical values (saturations) for the stimuli. Each participant was asked to judge the pairwise similarity between the stimuli from 1 (least similar) to 9 (most similar). The correlation was pretty high for the odd-numbered stimulus set ($r = .98$, $n = 10$) and the even-numbered stimulus set ($r = .99$, $n = 11$). Also, with the judged similarity as the dependent variable, the middle item was equally similar to the boundary items of the two categories for the odd-numbered stimulus set [$t(9) = -0.87$, $p =.41$] and so was for the even-numbered stimulus set [$t(10) = 0.32$, $p = .76$]. However, the left and right items to the middle one were more similar to the boundary items of the closer categories (all $p's < .01$). Thus, in no matter which stimulus set, the middle item was equally similar to either category. Therefore, the odd- and even-numbered participants' data were aggregated for data analysis.

Every participant went through four phases: training phase $\rightarrow$ review phase $\rightarrow$ recognition phase $\rightarrow$ transfer phase. The middle two correspond to the recognition task and the first and last phases correspond to the categorization task. In the training phase, on each trial in each of the 8 training blocks, the training item was presented on the center of computer monitor (in about 5° vision angle) until a response was made (by pressing key "s" or key "k" for one or the other category), followed by a corrective feedback (i.e., "correct" or "wrong").

In the review phase, in the Both condition, all exemplars were presented in random in 10 trials to the participants, who were asked to do nothing just memorize the circular colors for the latter recognition test. In the Low condition, only the 5 exemplars of the low-variability category were presented in random in 10 trials. Similarly, in the High condition, only the 5 high-variability category exemplars were presented in random in 10 trials. Thus, there were 10 trials in the review phase in no matter which condition. In the recognition phase, all 15 items shown in Figure 1 were presented one by one in random for all participants to judge whether it was previously seen in the review phase (by pressing key "y" for yes and key "n" for no). In the transfer phase, again, the same 15 items were presented in random to all participants for predicting the category of each item with no corrective feedback.

Results and Discussion

All participants data were included in data analysis. The results of data analysis are reported in the order of the experiment phases, with a pass given to the review phase, in which the participants were not asked to make any response.

Training Performance

The participants in no matter which condition could quickly learn the categories, with the learning accuracy increasing from 0.86 to 0.99 in the Both condition, from 0.82 to 0.99 in the High condition, and from 0.83 to 0.99 in the Low condition. A 3 (Condition) $\times$ 8 (Block) between-within subjects ANOVA shows a significant main effect of Block, $F(7,980) = 92.68$, $MSe = 0.004$, $p < .01$. However, there is no difference between the conditions, $F(2,140) < 1$ nor the interaction between Condition and Block, $F(14,980) < 1$.

Recognition Performance

The probability to be called as "old" for each of the 15 items is shown in Figure 3. For the ease of understanding the data, the stimulus order is reversed for the odd-numbered participants, making the high saturation values always correspond to the high-variability category.

Apparently, the participants did follow the instructions to do the recognition task. In the High condition (denoted by white circles), the probability of "old" response is extremely high for the exemplars of the high-variability category and drops dramatically for the exemplars of the low-variability category. In the Low condition (denoted by white diamonds), the pattern of the probability of "old" response is reversed. In the Both condition (denoted by black circles), the exemplars of both categories are judged as "old" with a probability larger than 0.75, above the 5 novel items. A 3 (Condition) $\times$ 15 (Item) between-within subjects ANOVA shows that the mean probability of "old" response is different in different conditions, $F(2,140) = 106.4$, $MSe = 0.23$, $p < .01$. Also, the probability of "old" response for different items is significantly different, $F(14,1960) = 56.37$, $MSe = 0.08$, $p < .01$.

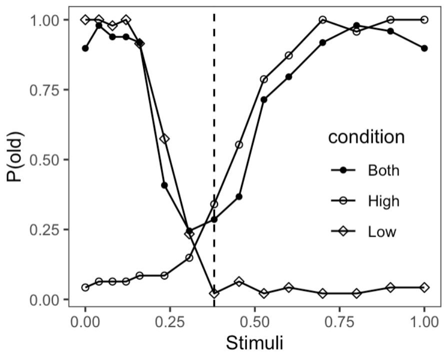

Figure 3: Recognition performance. The ordinate axis shows the probability of "old" response for each item and the abscissa axis represents the saturation values of the stimuli for even-numbered participants. The dashed vertical line marks the position of the middle item.

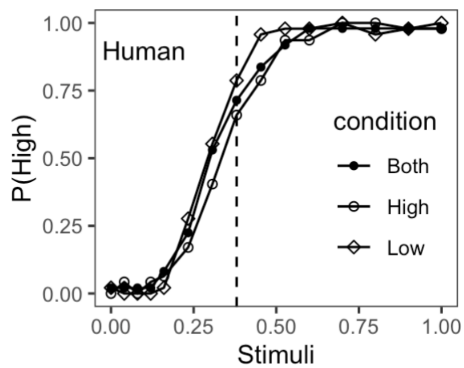

Figure 4: Transfer performance. The ordinate axis shows the probability of "High" response and the abscissa axis shows the transfer items. The Both condition, the High condition, and the Low condition are respectively denoted as black circles, white circles, and white diamonds. The dashed vertical line marks the position of the middle item.

The interaction between Condition and Item is also significant, $F(28,1960) = 110.10$, $MSe = 0.08$, $p < .01$.

It is worth checking how the participants would judge the middle item in the recognition task. Visual inspection of Figure 3 suggests that the middle item is more likely to be judged as "new" in all three conditions. However, this tendency differs across the conditions, $F(2,140) = 8.93$, $MSe = 0.15$, $p < .01$. This is because the probability of "old" response for it is significantly lower in the Low condition than the High condition, $F(1,140) = 15.56$, $MSe = 0.15$, $p < .01$. The reason for this result is unsure. Perhaps, the criterion for making an "old" judgment is higher in the Low condition. Also, it is worth examining whether the participants' recognition performance is correlated with the total similarity to the exemplars for all transfer items in this experiment. We will come back to this issue after introducing the results of computational modeling. Nonetheless, the manipulation over the exemplars to be rehearsed successfully alters the propensity to make an "old" response for the critical item.

Transfer Performance

Same as the recognition data, the transfer responses of the odd-numbered participants are rearranged to match the items observed by the even-numbered participants. The probability of "High" response for each transfer item is shown in Figure 4. The pattern of probability of "High" looks similar across these conditions. A 3 (Condition) $\times$ 15 (Item) between-within subjects ANOVA confirms this inspection that only the main effect of Item is significant, $F(14,1960) = 438.36$, $MSe = 0.06$, $p < .01$, the main effect of Condition is not significant, $F(2,140) = 1.56$, $MSe = 0.13$, $p = .21$, and there is no interaction effect between Condition and Item, $F(28,1960) = 1.13$, $MSe = 0.06$, $p = .29$.

Of our great interest is how the middle item would be classified. Visual inspection of Figure 4 shows that the middle item is more likely classified as the high-variability category in all conditions, implying the occurrence of CVE in all conditions. As every participant made only one response (i.e., "High" or "Low") for the middle item, the cumulation of "High" responses for it in each condition follows the binomial distribution. The null hypothesis that $H_o : p(High) = 0.5$ (and the scientific hypothesis that $H_1 : p(High) \neq 0.5$) is statistically tested with z test. The results rejects the null hypothesis in all conditions (for the Both condition, $z = 3.00$, $p < .01$; for the High condition, $z = 2.19$, $p < .05$; for the Low condition, $z = 3.94$, $p < .01$). The same results are obtained, if t test is conducted.

It is good for us to replicate CVE in this experiment. However, the null hypothesis significant testing cannot tell us the size of the true effect of CVE. Therefore, the true effect of H_1 over H_o in each condition is estimated via BF (Bayes Factor). The BFs are 20.15, 2.88, and 473.70 in the order of the Both condition, the High condition, and the Low condition. According to the suggestion of Kass and Raftery (1995), the evidence strength of the High condition is not worth more than a bare mention (BF < 3.2), whereas the strength of the evidence is strong for the Both condition (BF in between 10 and 100) and decisive for the Low condition (BF > 100). Therefore, it can be inferred that the Low condition induces the strongest CVE among all three conditions and the High condition the weakest. Our hypothesis is verified.

Computational Modeling

The SD-GCM is fit to each participant's transfer data individually with 4 free parameters: c, w_h (i.e., weighting for the high-variability category), w_l (i.e., weighting for the low-variability category), and γ, where $0 \leq c \leq 15$, $.001 \leq w_h \leq .999$, $.001 \leq w_l \leq .999$, and $0 \leq \gamma \leq 10$. The model predictions are shown in Figure 5.

It is of no doubt that the SD-GCM can accommodate well the observed data of every condition with the RMSD

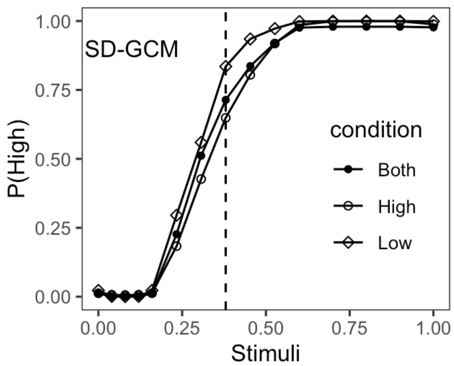

Figure 5: Predictions of SD-GCM for the transfer items.

(Root Mean Square Deviation) as 0.08 for the Both condition, 0.08 for the High condition, and 0.07 for the Low condition. The best-fit parameter values are list in Table 1. The high-variability category is generally less relied on for categorization than the low-variability category (i.e., $w_h < w_l$) with $t(48) = -6.65$, $p < .01$ for the Both condition, $t(46) = -5.32$, $p < .01$ for the High condition, and $t(46) = -13.06$, $p < .01$ for the Low condition. As CVE occurs in all conditions, this result supports our hypothesis that CVE tends to occur when focusing on the low-variability category for categorization.

Table 1: Statistics of best-fit parameter values in all conditions with mean in each entry and SD in parentheses.

	Both	High	Low
c	5.27 (3.78)	4.17 (2.47)	4.87 (3.06)
w_h	0.26 (0.40)	0.27 (0.41)	0.13 (0.30)
w_l	0.86 (0.32)	0.82 (0.36)	0.95 (0.20)
γ	9.95 (0.25)	9.83 (1.12)	9.95 (0.26)

Differential Reliance Degrees on High- and Low-Variability Categories

According to the results of the previous analysis, the difference between the weightings for the two categories $w_h - w_l$ can be treated as the index of the degree of CVE, the smaller the stronger. Considering the individual differences on CVE (see also Yang & Wu, 2014), the participants in each condition can be separated to two groups by whether or not s/he performed CVE (i.e., classifying the middle item as the high-variability category). A two-way between-subjects ANOVA shows that the CVE degree differs across conditions, $F(2,137) = 8.81$, $MSe = 0.11$, $p < .01$, and between the groups, $F(1,137) = 316.02$, $MSe = 0.11$, $p < .01$, and the CVE degree is influenced by the interaction effect between the conditions and the groups, $F(2,137) = 3.48$, $MSe = 0.11$, $p < .05$. See Table 2 for the CVE degrees in the two groups in all conditions.

Table 2: CVE degree $w_h - w_l$ in each group in each condition with sample size in parentheses.

	Both	High	Low
CVE	-0.94 (35)	-0.96 (31)	-1.00 (37)
non-CVE	0.24 (14)	0.27 (16)	-0.15 (10)

For the CVE group, the CVE degree is not different across the conditions, $F(2,100) < 1$, and the mean difference of $w_h - w_l$ is -.97, suggesting that these participants exclusively focused on the low-variability category for categorization. For the non-CVE group, the difference on the CVE degree is not significant across the conditions $F(2,37) = 1.93$, $MSe = 0.32$, $p = .16$ and the mean difference is .15, suggesting that the non-CVE group relied slightly more on the high-variability category than the low-variability category for categorization. To sum up, these results converge on the conclusion that CVE tends to occur when the low-variability category is largely focused for categorization.

Recognition and Categorization

Apparently, the recognition task successfully led the participants to focus on different categories for categorization. Thus, there should be some relation between the recognition performance (i.e., the probability of ”old” response) and the similarity to the exemplars (i.e., the weighted total similarity to the exemplars of the target category in the SD-GCM on fit to the categorization data). For the High condition, the larger the weighted total similarity to the high-variability category exemplars is, the more likely a transfer item is judged as ”old” [$r = .98$, $t(13) = 20.01$, $p < .01$]. Similarly, for the Low condition, the correlation between the recognition performance and the weighted total similarity to the low-variability category exemplars is extremely high [$r = .97$, $t(13) = 14.00$, $p < .01$]. However, there is no significant correlation in the Both condition between the recognition performance and the sum of weighted total similarity of the exemplars of the two categories [$r = .23$, $t(13) = 0.83$, $p = .42$]. This is because the novel items which are in between the exemplars of the two categories instead have a larger total similarity to all exemplars than the old items do. Nonetheless, our results can be viewed as a support for the argument that the recognition and categorization share the same memory system.

Conclusions

The goal of this study is to provide an exemplar account for the category variability effect, which has been regarded as a challenge for the exemplar model. A hypothesis was proposed that CVE tends to occur when the low-variability category is focused. The empirical and modeling results both support this hypothesis. Thus, CVE is no longer an effect that the exemplar model cannot account for.

There are a couple of theoretical and empirical implications of this study. First, this study shows that people can only rely on a single category to solve the problem of binary categorization. Although different from the conventional assumption in the exemplar model that categorization decision is made according to the difference between the evidence strengths for the candidate categories, with the summed dissimilarity to the exemplars of the target category as the response criterion, a single category can provide sufficient information for binary categorization. Second, apparently most of the participants preferred relying on the low-variability category for categorization when the two categories have unequal-sized variabilities. Although the SD-GCM can fit the data well, some more work is still needed for the SD-GCM to explain why people have such a preference. Third, the design of the present experiment is an innovation in experimental skills that embeds the recognition task in the categorization task to examine the relationships between them. The success of this design not only supports the relationships between recognition memory and categorization, but also provides an example for future studies with the attempt to examine the relationships between different cognitive functions.

References

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216–233.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in perception and categorization of multidimensional stimuli. *Journal of Experimental Psychological: Learning, Memory, and Cognition*, *14*, 33-53.

Austerweil, J. L., Liew, S., Conaway, N. B., & Kurtz, K. J. (2019). Creating something different: Similarity, contrast, and representativeness in categorization.

Cohen, A., Nosofsky, R., & Zaki, S. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*, 1165-1175.

Curtis, E. T., & Jamieson, R. K. (2018). Computational and empirical simulations of selective memory impairments: Coverging evidence for a single-system account of memory dissociation. *Quarterly Journal of Experimental Psychology*, *72*, 798–817.

De Schryver, M., Vandist, K., & Rosseel, Y. (2009). How many exemplars are used? Explorations with the Rex Leopold I model. *Psychonomic Bulletin & Review*, *16*(2), 337–343.

Hsu, A. S., & Griffiths, T. E. (2010, February). Effects of generative and discriminative learning on use of category variability. In S. Ohlsson & R. B. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society. (Available at SSRN: https://ssrn.com/abstract=1802162)

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *The Quarterly Journal of Experimental Psychology*, *62*, 550–575.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of American Statistical Association*, *90*(430), 773-795. doi: doi:10.2307/2291091

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

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 282–304.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of ”multiple-system” phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402.

Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *PNAS*, *109*, 333–338.

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

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*(4), 247–255.

Peirce, J. W. (2007). Psychopy-psychophysics software in python. *Journal of Neuroscience Methods*, *162*, 8-13.

Perlman, A., Hahn, U., Edwards, D. J., & Pothos, E. M. (2012). Further attempts to clarify the importance of category variability for categorisation. *Journal of Cognitive Psychology*, *24*(2), 203–220. doi: 10.1080/20445911.2011.61381

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), (pp. 21–59). New York, NY, US: Cambridge University Press.

Sakamoto, Y., Jones, M., & Love, B. C. (2008). Putting the psychology back into psychological models: Mechanistic versus rational approaches. *Memory & Cognition*, *36*(6), 1057–1065.

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

Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 893-907.

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

Yang, L.-X., & Wu, Y.-H. (2014). Category variability effect in category learning with auditory stimuli. *Frontiers in Psychology*, *5*, 1122. doi: 10.3389/fpsyg.2014.01122

Zaki, S. R., & Nosofsky, R. M. (2007). A high-distortion enhancement effect in the prototype-learning paradigm: dramatic effects of category learning during test. *Memory & Cognition*, *25*, 3088–3096.