Modeling the Category Variability Effect in an Exemplar-Similarity Framework

Florian I. Seitz (florian.seitz@unibas.ch)
Department of Psychology, Missionsstrasse 62A
Basel, Switzerland

Jana B. Jarecki (jana.jarecki@unibas.ch)
Department of Psychology, Missionsstrasse 62A
Basel, Switzerland

Jörg Rieskamp (joerg.rieskamp@unibas.ch)
Department of Psychology, Missionsstrasse 62A
Basel, Switzerland

Abstract

The category variability effect describes assigning objects to high-variability categories. We show that similarity-based categorization theories can predict the category variability effect and conduct a rigorous empirical test. In an optimized categorization experiment, participants learned to assign geometrical figures to a high-variability and a low-variability category and then categorized transfer stimuli located between the categories. We compared a formal model that ignores category variability (Euclidean model) to one that considers category variability (Mahalanobis model) during similarity computation. The data ($N = 43$) revealed that most participants did not show the category variability effect, in line with the Euclidean model. Nevertheless, the Mahalanobis model consistently described the participants that selected the high-variability category. This demonstrates that—contrary to previous claims—similarity can explain the category variability effect. However, in our data, most people do not seem to show the effect, maybe because the low-variability category was more coherent than the high-variability category.

Keywords: Similarity; Perceptual categorization; Category variability effect; Exemplar theory; Computational modeling

Introduction

Different categories have different properties. Whereas some categories include only a narrow range of objects, other categories encompass objects whose features vary much more. Body weight, for example, varies more among males than females in some species: Male Great Danes can weigh 54 to 91 kg (a large variability), but female Great Danes weigh only 45 to 59 kg (a small variability). Hereafter, category variability refers to the feature variances across the members of a category. This work examines to what extent people make use of the variability of categories to categorize new objects. To this end, we test how people categorize objects lying between two categories with differing variability (see Fig. 1 for a schematic illustration). The novelty of our approach is twofold: First, our experimental design was mathematically optimized in simulations to study category variability (optimal experimental design, Myung & Pitt, 2009). Second, we developed a cognitive model to test a similarity-based categorization mechanism (Nosofsky, 1989) that processes the category variability during similarity computation, thereby predicting the category variability effect defined below.

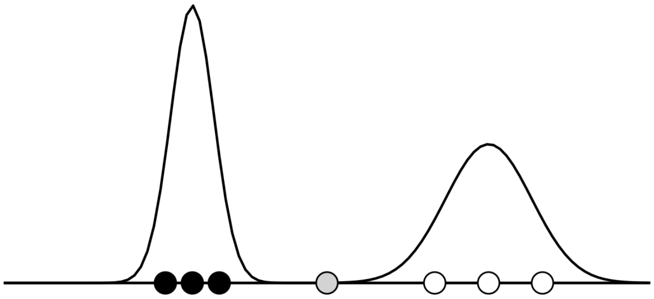

Figure 1: *Category variability*. Shown are a category with a small variability concerning a feature (e.g., body weight) and a category with a large variability concerning this feature. Assigning the new object lying between the categories (shown as grey ”o”) to the high-variability category represents the category variability effect.

The Category Variability Effect in Categorization

The *category variability effect* refers to people assigning objects to high-variability categories rather than to comparably low-variability categories. Most research investigating the effect of category variability on human categorizations has tested how people categorize objects that are located between a high-variability and a low-variability category (cf. Rips, 1989 and Cohen, Nosofsky, & Zaki, 2001). Fig. 1 illustrates this for a single-feature case, in which a to-be-classified object (shown as a grey ”o”) could belong to a category containing members that vary a lot with respect to this feature (white ”o”s) or to a category that varies less with respect to this feature (black ”o”s). According to the category variability effect, the object is systematically assigned to the high-variability category. Importantly, categorization research has shown that some people display the category variability effect, whereas others do not (Cohen et al., 2001; Fried & Holyoak, 1984; Hsu & Griffiths, 2010; Perlman, Hahn, Edwards, & Pothos, 2012; Rips, 1989; Sakamoto, Jones, & Love, 2008; Stewart & Chater, 2002; Yang & Huang, 2021; Yang & Wu, 2014).

In Cohen et al. (2001), participants selected the high-variability category in 63% of cases in one experiment in-

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

volving objects with two features, but in only 30% to 47% of cases in the other experiment with single-feature objects. Similarly, Stewart and Chater (2002) found that participants mostly assigned a mid-point object to the low-variability category. Only when the salience of the category variability was increased by showing the category members simultaneously or by providing an explicit hint about the difference in variability between categories, the category variability effect occurred. In contrast, the results of Perlman et al. (2012) and Yang and Huang (2021) suggest that about 75% of participants show the category variability effect, which only disappears when the differences in category variability are made less salient. In Yang and Wu (2014), two equally sized subsamples of participants (about 28% each) consistently chose the high-variability or the low-variability category, respectively (the remaining participants' classifications depended on the preceding object). This highlights large interindividual differences concerning the category variability effect.

Explaining the Category Variability Effect With Similarity-Based Categorization Theories

The category variability effect has been considered evidence against the theory assuming that people categorize objects based on their similarity to previously experienced category members (*exemplars*; Smith & Sloman, 1994; Yang & Wu, 2014). According to this *exemplar-similarity theory*, an object is assigned to the category that contains the exemplars to which the object is most similar (e.g., Nosofsky, 1986). A category becomes more probable as the object's similarity to the category's exemplars increases. Accordingly, similar objects are assigned to the same category and dissimilar objects to different categories—an assumption that is well-supported empirically (Kruschke, 2008; Nosofsky, 1986, 1989, 2011).

In the context of category variability, an object located in the middle between two categories (the grey ”o” in Fig. 1) is geometrically closer and therefore more similar to the members of the small-variability category than to the members of the high-variability category. Therefore, according to similarity, the object should be assigned to the low-variability category; yet, the category variability effect means selecting the high-variability category. Consequently, it has been argued that similarity-based categorization mechanisms such as the generalized context model cannot produce a category variability effect (Smith & Sloman, 1994; Yang & Wu, 2014; but see Nosofsky & Johansen, 2000 and Yang & Huang, 2021).

In the following, we show that psychological similarity can account for the category variability effect. Cognitive systems can process similarity in various ways, and one frequently-used psychological similarity—the *Euclidean similarity* (e.g., Nosofsky, 1987, 2011)—ignores the category variability and thus cannot predict the category variability effect. Another way to compute similarity, rarely used in cognitive research (except Battleday, Peterson, & Griffiths, 2020) compared to machine learning research (e.g., Weinberger & Saul, 2009), considers the category variability and can predict the category variability effect: the *Mahalanobis similarity*. This work compares these two similarities in the exemplar-similarity framework to test if people make use of the variability of categories to categorize objects and thereby display the category variability effect (see below for a formal implementation).

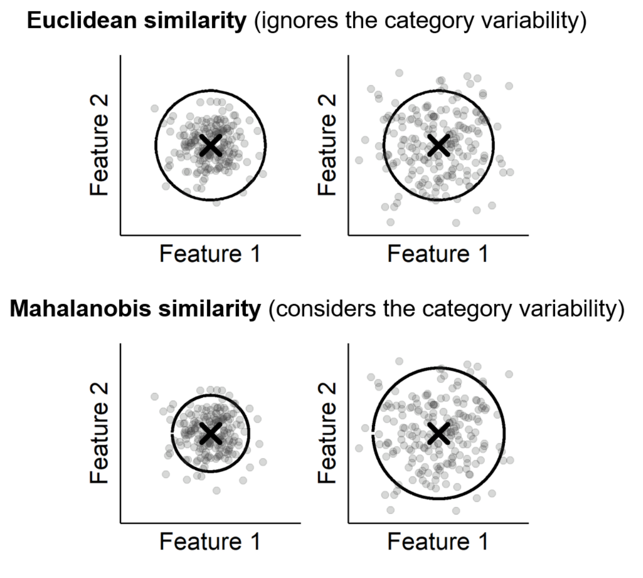
Euclidean similarity (ignores the category variability)
Feature 2
Feature 1
Feature 2
Feature 1
Mahalanobis similarity (considers the category variability)
Feature 2
Feature 1
Feature 2
Feature 1

Figure 2: *Similarity and category variability*. Shown are the locations in a two-feature space with a constant similarity to the center (”x”) as measured by the Euclidean similarity (top) or the Mahalanobis similarity (bottom), given categories (grey points) with different variability in their features.

Fig. 2 illustrates how category variability can affect psychological similarity. The grey circles in the figure show a low-variability category and a high-variability category in a two-feature space, and the solid lines depict the locations that have the same Euclidean similarity or the same Mahalanobis similarity to the center ”x”. Importantly, the category variability does not affect which locations have a constant Euclidean similarity to the center, as this similarity ignores the category variability and implicitly assumes unit feature variances for all categories. Unlike this, the locations around the center with a constant Mahalanobis similarity adapt to the category variability and increase with larger feature variances. Thus, the Mahalanobis similarity between a to-be-classified object and an exemplar increases with larger feature variances in the category to which the exemplar belongs. For the category variability effect, this entails that even though an object lying between two categories is geometrically closer to the low-variability category, it has the larger Mahalanobis similarity to the high-variability category, because the categories' difference in variability quickly outweighs their difference in geometrical closeness to the to-be-classified object.

Formal Model Implementation

We formalized the Euclidean similarity and the Mahalanobis similarity within the generalized context model (see Nosofsky, 1989), which assumes that the similarity to all previously experienced exemplars determines categorization. Given two categories labeled A and B, the model computes the evidence that an object belongs to category A from the summed similarity to all category A exemplars relative to the summed similarity to all exemplars in both categories. Formally, the evidence $E(A|i)$ of object i for category A is

$$E(A|i) = \frac{\sum\limits_{j \in A} s_{ij}}{\sum\limits_{j \in A} s_{ij} + \sum\limits_{k \in B} s_{ik}}, \tag{1}$$

where s_{ij} is the psychological similarity between object i and exemplar j from category A. In the following, we describe two model versions that either ignore or consider the category variability during similarity computation.

Ignoring Category Variability: Euclidean Model. One version of the generalized context model applies the Euclidean similarity, which ignores category variability. This means that a category's feature variances do not affect how similar the category's exemplars are to another object. Specifically, the Euclidean similarity s_{ij} of object i to exemplar j is

$$s_{ij} = \exp\left[-c \cdot (\mathbf{W}(\mathbf{x}_i - \mathbf{x}_j))^\top (\mathbf{x}_i - \mathbf{x}_j)\right], \tag{2}$$

where the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ contain the N feature values of i and j, respectively[1]. The equation has N free model parameters: The N-by-N diagonal matrix $\mathbf{W} = \text{diag}(w_1, \ldots, w_N)$ (with $0 \leq w_n \leq 1$ for all $n \in \{1, \ldots, N\}$ and $\sum_{n=1}^{N} w_n = 1$) contains individual people's attention weights for the N features, and the scalar c (with $c > 0$) describes how steeply the similarity declines with higher feature value differences $(\mathbf{x}_i - \mathbf{x}_j)$.

Considering Category Variability: Mahalanobis Model. Another, new version of the generalized context model uses the Mahalanobis similarity, which considers the category variability. This means that a category's feature variances affect how similar the category's exemplars are to another object. Larger feature variances (i.e., a greater category variability) increase the similarity between the category's exemplars and another object. The Mahalanobis similarity $s_{ij \in A}$ of object i to exemplar j from category A is computed as

$$s_{ij \in A} = \exp\left[-c \cdot (\mathbf{W}(\mathbf{x}_i - \mathbf{x}_j))^\top \mathbf{K}_A^{-1} (\mathbf{W}(\mathbf{x}_i - \mathbf{x}_j))\right], \tag{3}$$

where $\mathbf{K}_A^{-1}$ is the inverse of the N-by-N variance-covariance matrix of category A. This matrix contains in its main diagonal the N feature variances, computed across all previously experienced exemplars of category A. An analogous matrix $\mathbf{K}_B^{-1}$ is used to compute the Mahalanobis similarity to category B exemplars. Eq. 3 is based on Mahalanobis (1936).

Relating the Two Models. The Euclidean similarity and the Mahalanobis similarity are both based on squared feature value differences and are thus related to each other. However, the Mahalanobis similarity extends the Euclidean similarity by standardizing the feature value differences by the category variability (see the category-specific variance-covariance matrix in Eq. 3). The predictions of the two similarities converge as the feature variances within categories approach 1 (given uncorrelated features within categories). Any other category variability only affects the Mahalanobis similarity, which increases with higher category variability, but not the Euclidean similarity, which remains constant (see also Fig. 2).

Modeling People's Category Responses. To model participants' responses, we transformed the category evidence to category response probabilities using the softmax choice rule

$$\Pr(A|i) = \frac{\exp(\tau \cdot E(A|i))}{\exp(\tau \cdot E(A|i)) + \exp(\tau \cdot E(B|i))} \tag{4}$$

with the free parameter τ ($\tau > 0$, a higher value means more deterministic responding) and $E(A|i)$ as computed in Eq. 1. Choice rule parameters such as τ can correlate negatively with parameter c (e.g., Krefeld-Schwalb, Pachur, & Scheibehenne, 2022). We accepted this dependence as we did not focus on participants' parameter estimates but on modeling their categorizations, which are often more deterministic than predicted without a choice rule (Krefeld-Schwalb et al., 2022).

Summary We adapted the psychological similarity function of the generalized context model to investigate the category variability effect. The model can categorize new objects by ignoring the category variability (the original Euclidean model) or by considering the category variability (the new Mahalanobis model). Importantly, according to the Mahalanobis model, similarity increases with greater category variability, leading to the category variability effect (for a related approach using the standard deviation of the categories' features in a prototype model, see Sakamoto et al., 2008).

Experiment

To investigate the category variability effect and model it in the exemplar-similarity framework described above, we conducted a binary categorization experiment in which participants categorized objects lying between a category with small variability and a category with large variability.

Participants. We aimed for 42 participants, which allows testing if one of the cognitive models describes more than half the participants, given $\alpha = .05$, $1 - \beta = .95$, and $g = 0.25$ as smallest effect size of interest (justified by the optimal experimental design detailed below). In total, 49 participants, recruited over Prolific Academic (`www.prolific.co`), participated in an online study and received a compensation of £8.72 (about \$12 during the study period). Six participants were excluded for not having reached the category learning accuracy criterion (see Procedure), resulting in a final sample of $N = 43$ (8 females, $M_{age} = 27.47$ years, $SD_{age} = 7.94$ years, age range: 18-56 years). The experiment lasted about 40 minutes and was approved by the ethics board of the psychology department of the University of Basel (#025-18-6).

[1] Non-scalar variables (vectors and matrices) are shown in **bold**.

Optimal Experimental Task Design. The experiment contained a trial-by-trial supervised category learning task followed by an unsupervised transfer task (e.g., Nosofsky, 1989). Participants learned by feedback to classify stimuli with two continuous features into two categories (the category structure) and then classified transfer stimuli (new feature value combinations) without feedback. We used a simulation-based optimal experimental design procedure (as in Myung & Pitt, 2009) to find transfer stimuli that maximally discriminate between the predictions of the Euclidean model and the Mahalanobis model, given training on the category structure from the learning task. Our optimization procedure aimed to ensure that both models can learn the category structure but make maximally opposite category predictions for the transfer stimuli afterward. This allows for a fair test of the influence of category variability in the transfer task without favoring any model during category learning.

In the simulations, the two models were trained to learn the category labels of several feature value combinations (the category structure) and then predicted the category labels of new feature value combinations (the transfer stimuli). The simulations encompassed a range of category structures, transfer stimuli, and admissible model parameters, constraining that in all category structures, the two features varied more in one of the two categories and were uncorrelated and that both categories had the same average feature value combination (centroid). The optimized task design resulting from these simulations (i.e., the category structure and the transfer stimuli T1 to T8 that maximally differentiate between the Euclidean and Mahalanobis models after learning) is shown in Fig. 3. The figure also shows some less well-discriminating filler stimuli.

In this design, the aggregate model predictions are the same for the different transfer stimuli $\mathrm{T} \in \{\mathrm{T1}, ..., \mathrm{T8}\}$. The Euclidean model, which ignores category variability, assigns all transfer stimuli to the low-variability category *A* with $\mathrm{Pr}_{\mathrm{Eucl}}(A \mid \mathrm{T}) = .74$, because the stimuli differ less from the category *A* exemplars than from the category *B* exemplars. In turn, the Mahalanobis model, which considers category variability, assigns the transfer stimuli to the high-variability category *B*, $\mathrm{Pr}_{\mathrm{Maha}}(B \mid \mathrm{T}) = .72$, as this category occupies a larger area of the feature space, which outweighs the geometrical closeness of the transfer stimuli to category *A*. In contrast to past studies (e.g., Hsu & Griffiths, 2010; Yang & Wu, 2014), this design cannot be learned by unidimensional rules and might thus foster the use of similarity-based categorization strategies (Rouder & Ratcliff, 2006). The category structure can be learned by separating ”moderate” from ”extreme” exemplars; however, such a rule arguably implicitly considers the smaller category variability of the ”moderate” category *A* relative to the ”extreme” category *B*.

Materials. The experiment was programmed in PsychoPy3 (Peirce, 2007). Participants classified geometric shapes whose features were a circle of varying size and a line of varying orientation (similar to Nosofsky, 1989; for a visualization, see Fig. 3). Participants classified the stimuli with the ”e” and ”i” keys, and the assignment of the keys (e and i) to the categories (*A* and *B*) was randomized across participants.

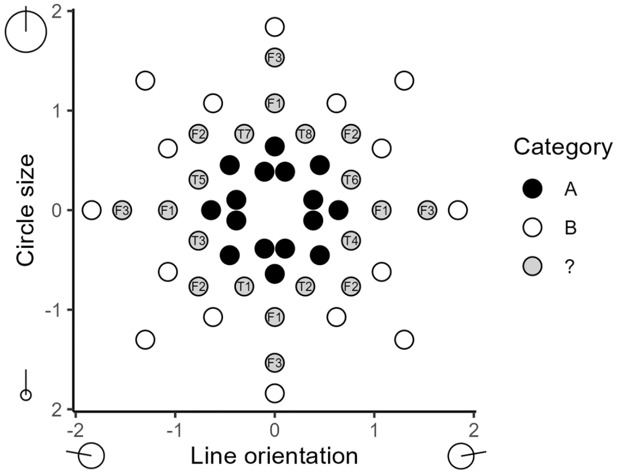

Figure 3: *Optimal experimental design.* Each point is a stimulus, with the color denoting the category. Grey points are the stimuli without a category label: the optimally-discriminating transfer stimuli (T1 to T8) and exploratory, less well-discriminating filler stimuli (grouped as F1 to F3).

Procedure. Participants' task was to assign stimuli with two features to one of two categories (similar to Nosofsky, 1989). After familiarizing themselves with possible feature value combinations, participants repeatedly categorized stimuli one at a time until they learned the category structure in Fig. 3. In each trial, they saw a stimulus (a combination of two feature values), categorized it by pressing a key, and got visual feedback about their categorization in the form of a smiley and a notification for 1000 ms. To ensure that participants learned the category structure equally well, the learning phase ended after a participant correctly classified more than 90% of the last 100 trials (similar to Seitz, von Helversen, Albrecht, Rieskamp, & Jarecki, 2023). If a participant did not reach this accuracy criterion in 1,280 trials (40 blocks), learning ended and the respective participant was excluded. To encourage learning, participants received a performance message every 50 trials starting at trial 100, indicating how many of the last 100 stimuli they classified correctly. After learning, participants classified the transfer stimuli (denoted by Ts in Fig. 3) without feedback. Participants were informed of the absence of feedback and categorized the eight transfer stimuli three times, the 32 old learning stimuli twice, and 12 filler stimuli (denoted by Fs in Fig. 3 and grouped into three groups based on their location in feature space), resulting in 100 transfer trials with randomized order.

Results

Analyses were conducted in R (v3.6.1, R Core Team, 2017). Inferential statistics used the *lme4* package (v1.1-23, Bates, Mächler, Bolker, & Walker, 2015); cognitive modeling used the *cognitivemodels* package (v0.0.12, Jarecki & Seitz, 2020).

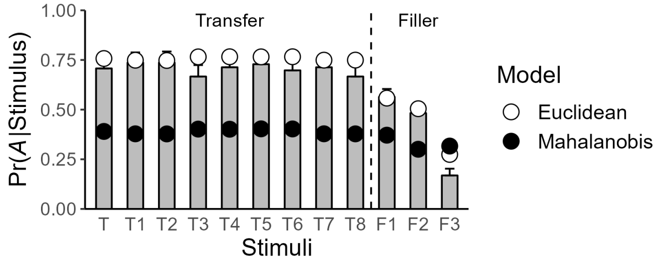

Figure 4: *Results.* Bars and whiskers show the mean and standard error, respectively, of how participants categorized the transfer stimuli (left; the leftmost bar averages across transfer stimuli) and the filler stimuli (right). Shapes show the mean model predictions across participants (see text for details).

Category Learning. Participants reached the 90% learning accuracy criterion after on average M = 420 learning trials (Mdn = 350, SD = 226, range = 132-1019). The accuracy for the learning stimuli remained high in the transfer phase (M = 89%, Mdn = 89%, SD = 5 percentage points; the values are computed across each participant's mean accuracy).

Category Variability Effect. We tested if participants show the category variability effect and assign the transfer stimuli to the high-variability category B. Alternatively, participants can ignore the category variability and assign the transfer stimuli to the low-variability category A. Consistent with ignoring category variability, participants assigned the transfer stimuli to category A in M = 71% of cases (see Fig. 4).

A linear mixed model with logit link modeling participants' category responses as a function of the transfer stimuli (fixed effect with eight levels) with a participant-wise random intercept corroborated these results, see Table 1 for the resulting log-odds regression coefficients. The intercept coefficient indicates that participants assigned the transfer stimuli to category A with an average probability of .75, $p < .001$. The fixed effect coefficients show that participants categorized the individual transfer stimuli similarly to the grand mean (range: .70-.78, e.g., $\Pr(A|\mathrm{T1}) = .75 + .03 = .78$, $ps \geq .23$). Accordingly, the fixed effect coefficients do not differ significantly from each other, all ps = 1, based on Holm-Bonferroni corrected post-hoc contrasts (Holm, 1979).

Participants thus assigned the transfer stimuli predominantly to the low-variability category A. These aggregate results are at odds with the category variability effect (see also Cohen et al., 2001 and Yang & Wu, 2014) and do not seem to reflect a pure category A bias as participants categorized the three groups of filler stimuli much more variably (see Fig. 4).

Computational Modeling and Model Comparison. We applied computational cognitive modeling to gain more insight into how the cognitive processes differ across participants. The free parameters of the Euclidean and Mahalanobis models (i.e., two attention weights ws summing up to 1, sensitivity c with $0 < c \leq 10$, and temperature τ with $0.1 \leq \tau \leq 10$) were estimated with maximum likelihood from individual participants' learning phase data without the first block. Table 2 shows the resulting parameter estimates.

Table 1: Fixed effect estimates from a linear mixed model with logit link on the category A choices.

Stimulus	Coefficient	Pr(A)	SE	z	p
Intercept	1.10	.75	0.18	6.13	< .001
T1	0.17	+.03	0.20	0.83	.40
T2	0.17	+.03	0.20	0.83	.40
T3	−0.23	−.04	0.19	−1.19	.23
T4	0.03	+.01	0.20	0.15	.88
T5	0.12	+.02	0.20	0.61	.54
T6	−0.06	+.01	0.19	−0.30	.77
T7	0.03	+.01	0.20	0.15	.88
T8	−0.23	−.04	0.19	−1.19	.23

Note. The model was run with sum-to-zero contrasts. This means the intercept is the grand mean across transfer stimuli, and the fixed effects are the differences between the stimulus mean and the intercept (Singmann & Kellen, 2017).

Table 2: Parameter estimates resulting from fitting the model to the learning phase data with maximum likelihood.

	Euclidean model			Mahalanobis model		
Parameter	M	Mdn	SD	M	Mdn	SD
w_{size}	.65	.75	.28	.64	.63	.32
w_{angle}	.35	.25	.28	.36	.37	.32
c	7.42	7.22	2.51	2.40	0.14	4.18
τ	2.41	2.37	0.80	5.62	3.90	5.24

Note. Parameter estimates are aggregated across participants. Note that c and τ negatively correlate ($r = -.46$) and may trade off against each other (Krefeld-Schwalb et al., 2022).

To test whether people consider category variability during categorization, we conducted a model comparison on the eight transfer stimuli (which were not used for model fitting, i.e., the hold-out data). Individual participants' best-fitting parameter estimates were used to predict their categorizations for the transfer stimuli. Based on the models' predictions and the participants' responses, the log-likelihoods of the models were computed for each participant. These log-likelihoods were further transformed into model evidence strengths (measured as Akaike weights, $w(\mathrm{AIC})$, Wagenmakers & Farrell, 2004), which range from 0 to 1 and which were used for the model comparison. At the aggregate level, the mean model evidence strength across participants was used. At the individual level, each participant was assigned to the model with the highest evidence strength $w(\mathrm{AIC})$ (given $w(\mathrm{AIC}) \geq .67$), resulting in the number of participants each model can describe. In addition to the Euclidean and the Mahalanobis models, the model comparisons included a baseline random choice model, predicting categorizations of $\Pr(A|\mathrm{T}) = .50$.

Table 3: Descriptive model fit measures across all participants and across the participants assigned to the respective model.

	All participants				Assigned participants				
Model	w(AIC)	ℓ	Accuracy	*MSE*	n (%)	w(AIC)	ℓ	Accuracy	*MSE*
Euclidean	.65	-13.40	71%	.19	27 (63%)	.96	-10.48	82%	.14
Mahalanobis	.20	-18.93	33%	.29	7 (16%)	.88	-12.97	66%	.18
Random choice	.15	-16.64	50%	.25	2 (5%)	.89	-16.64	50%	.25

Note. The fit measures are: w(AIC) = mean evidence strength (Akaike weights), ℓ = mean log-likelihood, Accuracy = mean correspondence with participants' responses when assigning stimuli to the most likely category. *MSE* = mean squared error.

At the aggregate level, the Euclidean model clearly outperforms the Mahalanobis model, suggesting that participants on average ignored the category variability when categorizing the transfer stimuli. Specifically, the Euclidean model correctly predicted 71% of the transfer categorizations, while the Mahalanobis model predicted only 33%. Both accuracies differ significantly from the 50% accuracy of the random choice model (Euclidean model: $\chi^2(1)$ = 176.68, $p < .001$; Mahalanobis model: $\chi^2(1)$ = 124.88, $p < .001$). Also in terms of model evidence, the Euclidean model excels with a mean Akaike weight w(AIC) = .65, which exceeds the evidence for the Mahalanobis model (w(AIC) = .20) by a factor of 3.27 and the evidence for the random choice model (w(AIC) = .15) by a factor of 4.32. Table 3 summarizes the model fit indices.

Also at the individual level, the Euclidean model excels (see Fig. 5). It describes n = 27 participants (63%; among them n = 18 with decisive evidence[2]). Still, the Mahalanobis model also describes n = 7 participants (16%; n = 2 with decisive evidence). For most of the remaining 9 participants, there was no clear winning model (n = 7, 16%); the random choice model described the other n = 2 (5%). The number of described participants varied substantially across models (exact multinomial test, $p < .001$), with the Euclidean model outperforming the Mahalanobis model, $p = .002$, and the random choice model, $p < .001$. Still, the Euclidean model did not describe more than half the participants (exact binomial test, $p = .13$), pointing up the interindividual differences.

The 27 participants described by the Euclidean model selected the low-variability category in 83% of cases (*Mdn* = 83%, *SD* = 13 percentage points). Also, the Euclidean model correctly predicted 82% of these 27 participants' choices, which significantly exceeds the accuracy of 50% resulting from a random choice model, $\chi^2(1) = 268.35$, $p < .001$. In turn, the 7 participants described by the Mahalanobis model selected the high-variability category in 55% of cases (*Mdn* = 58%, *SD* = 7 percentage points), and the model correctly predicted 66% of their choices, which is also significantly larger than 50%, $\chi^2(1) = 16.72$, $p < .001$. This shows that—contrary to previous claims (e.g., Smith & Sloman, 1994)—psychological similarity can account for the category variability effect; yet, in our data, few people show the effect.

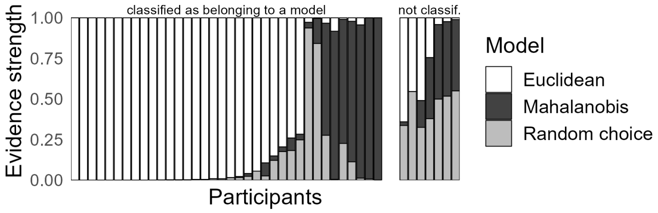

Figure 5: *Evidence strength for the cognitive models.* Each bar represents a participant and shows the proportion of evidence for each model (the stacked total evidence sums to 1).

Discussion

This work investigated to what extent people make use of the variability of categories to categorize objects. Within the exemplar-similarity framework, we compared the Euclidean model, which formalizes ignoring category variability, against a new Mahalanobis model, which formalizes considering category variability. Inferential statistics and predictive cognitive modeling strongly suggest that the categorizations of many participants (n = 27, 63%) ignored the category variability and followed the predictions of the Euclidean model. Only a few participants (n = 7, 16%) consistently behaved in line with the Mahalanobis model and seemed to show the category variability effect (see Yang & Wu, 2014).

One reason for this might be the larger coherence of the low-variability category *A* relative to category *B* (see Fig. 3), which might have led participants to assign the new stimuli to category *A*, unless they were very untypical of it (close to an A/not-A task, see Casale & Ashby, 2008). We believe our category structure is suited to study the effect of category variability in the exemplar-similarity framework, as it cannot be learned by simple rules, thereby arguably fostering the use of similarity-based strategies. Moreover, we found similar results in another study using an information-integration category structure with two psychologically meaningful categories (Seitz, Jarecki, & Rieskamp, 2021).

Taken together, our work shows that—contrary to previous claims (e.g., Smith & Sloman, 1994)—similarity can process category variability and adds to the literature that extends the exemplar-similarity theory to explain the category variability effect (Nosofsky & Johansen, 2000; Yang & Huang, 2021).

[2] As in Bayes factor interpretation, a model's evidence is decisive if it is at least 100 times larger than the next-best model's evidence.

Acknowledgements

This work was supported by the Swiss National Science Foundation [grant number P0BSP1_195389 to F. I. Seitz].

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Battleday, R. M., Peterson, J. C., & Griffiths, T. L. (2020). Capturing human categorization of natural images by combining deep networks and cognitive models. *Nature Communications*, *11*(1), 1–14. doi: 10.1038/s41467-020-18946-z

Casale, M. B., & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Perception & Psychophysics*, *70*(6), 983–999. doi: 10.3758/pp.70.6.983

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*(8), 1165–1175. doi: 10.3758/BF03206386

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 234–257. doi: 10.1037//0278-7393.10.2.234

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. doi: 10.2307/4615733

Hsu, A. S., & Griffiths, T. E. (2010). Effects of generative and discriminative learning on use of category variability. In *32nd Annual Conference of the Cognitive Science Society.*

Jarecki, J. B., & Seitz, F. I. (2020). Cognitivemodels: An R package for formal cognitive modeling. In T. C. Stewart (Ed.), *Proceedings of the 18th International Conference on Cognitive Modelling* (pp. 100–106). University Park, PA: Applied Cognitive Science Lab, Penn State.

Krefeld-Schwalb, A., Pachur, T., & Scheibehenne, B. (2022). Structural parameter interdependencies in computational models of cognition. *Psychological Review*, *129*(2), 313–339. doi: 10.1037/rev0000285

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 267–301). New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9780511816772

Mahalanobis, P. C. (1936). On the generalized distance in statistics. In *Proceedings of the National Institute of Science of India* (Vol. 2, pp. 49–55).

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*(3), 499–518. doi: 10.1037/a0016104

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 87–108. doi: 10.1037/0278-7393.13.1.87

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, *45*(4), 279–290. doi: 10.3758/BF03204942

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In *Formal approaches in categorization* (pp. 18–39). New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9780511921322

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402. doi: 10.1007/BF03543066

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi: 10.1016/j.jneumeth.2006.11.017

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: 10.1080/20445911.2011.613818

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge University Press. doi: 10.1017/CBO9780511529863.004

Rouder, J. N., & Ratcliff, R. (2006). Comparing exemplar- and rule-based theories of categorization. *Current Directions in Psychological Science*, *15*(1), 9–13. doi: 10.1111/j.0963-7214.2006.00397.x

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. doi: 10.3758/MC.36.6.1057

Seitz, F. I., Jarecki, J. B., & Rieskamp, J. (2021). People are insensitive to within-category feature correlations in categorization. In T. C. Stewart (Ed.), *Proceedings of the 19th International Conference on Cognitive Modelling* (pp. 255–256). University Park, PA: Applied Cognitive Science Lab, Penn State.

Seitz, F. I., von Helversen, B., Albrecht, R., Rieskamp, J., & Jarecki, J. B. (2023). Testing three coping strategies for time pressure in categorizations and similarity judgments. *Cognition*, *233*, 105358. doi: 10.1016/j.cognition.2022.105358

Singmann, H., & Kellen, D. (2017). An introduction to mixed models for experimental psychology. In D. H. Spieler &

E. Schumacher (Eds.), *New methods in neuroscience and cognitive psychology* (pp. 4–31). Psychology Press.

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & Cognition*, *22*(4), 377–386. doi: 10.3758/BF03200864

Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 893–907. doi: 10.1037/0278-7393.28.5.893

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. doi: 10.3758/BF03206482

Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, *10*(2), 207–244. doi: 10.5555/1577069.1577078

Yang, L.-X., & Huang, T.-L. (2021). Exemplar account for category variability effect: Single category based categorization. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, pp. 2699–2705).

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