

# People are Insensitive to Within-Category Feature Correlations in Categorization

Florian I. Seitz (florian.seitz@unibas.ch)

Department of Psychology, Missionsstrasse 62A  
Basel, 4055 Switzerland

Jana B. Jarecki (jana.jarecki@unibas.ch)

Department of Psychology, Missionsstrasse 62A  
Basel, 4055 Switzerland

Jörg Rieskamp (joerg.rieskamp@unibas.ch)

Department of Psychology, Missionsstrasse 62A  
Basel, 4055 Switzerland

**Keywords:** categorization; similarity; cognitive modeling; correlational sensitivity; within-category information

## Introduction

This work investigates people’s sensitivity to within-category feature correlations in perceptual categorization by testing two types of psychological similarity against each other.

Similarity is a central component of many psychological categorization theories. Exemplar theories, for instance, assume that people categorize new objects based on their similarity to previously seen category members (Nosofsky, 1986, 1989). Traditionally, the underlying psychological similarity between a pair of objects is modeled as the sum of the objects’ squared feature value differences (*Euclidean similarity*; e.g., Goldstone, 1994). The Euclidean similarity, however, ignores the distribution of objects by assuming uncorrelated features within categories. In turn, the *Mahalanobis similarity* extends the Euclidean similarity by accounting for within-category feature correlations. Results from machine learning have shown that the Mahalanobis similarity can outperform the Euclidean similarity in categorization problems involving correlated features within categories (Mao & Jain, 1996; Weinberger & Saul, 2009). Yet, in the psychological categorization literature, there are mixed results regarding the extent to which people take within-category feature correlations into account (Chin-Parker & Ross, 2002; Ell, Smith, Peralta, & Hélie, 2017; Lancaster, Shelhamer, & Homa, 2013).

Therefore, the present work investigated if people use within-category feature correlations for categorization. Our work rigorously compared the correlation-insensitive Euclidean similarity against the correlation-sensitive Mahalanobis similarity by means of mathematical modeling on data from an optimized category learning task.

## Methods

We designed a standard trial-by-trial supervised, binary category learning task (e.g., Nosofsky, 1989) with two strongly correlated features within each category ( $r = .98$ ). In the task, participants learned to categorize a set of stimuli from feedback and then categorized new test stimuli

without feedback. The category structure was selected using simulation-based optimal experimental design (Myung & Pitt, 2009). Our formal modeling framework were two versions of the exemplar model of Nosofsky (1986, i.e., the *generalized context model*); one version used the Euclidean similarity, the other used the Mahalanobis similarity.

To optimize the experimental design, we searched for a category structure that both model versions can learn accurately and that maximizes the classification prediction differences between the two model versions for the test stimuli. Figure 1 shows that in the resulting optimal design the Euclidean similarity assigns the test stimuli into the category with lower straight-line distances, whereas the Mahalanobis similarity assigns them into the category with the matching correlational structure.

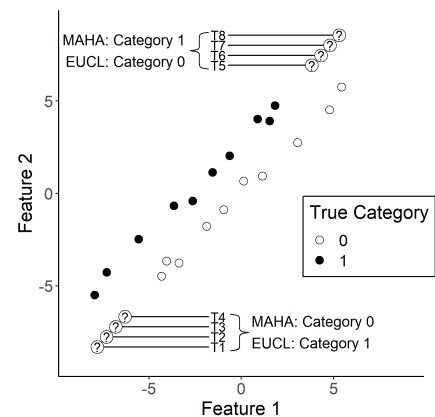


Figure 1: *Category structure of the task involving a high within-category feature correlation.* Shown are the 20 learning stimuli with their true category (black and white circles) and the 8 test stimuli presented after learning without feedback (circles with question mark). EUCL = predictions of the model using the Euclidean similarity; MAHA = predictions of the model using the Mahalanobis similarity.

In the experiment, participants ( $N = 43$ ; 14 females,  $M_{age} = 25.56$  years,  $SD_{age} = 6.81$  years) learned to classify 20 stimuli with feedback until they were more than 90% accurate

across the last 100 trials; then, they classified 8 new test stimuli without feedback. The stimuli were geometric figures consisting of a circle of varying size and a line of varying orientation (as in Nosofsky, 1989); category labels and visual feature assignments were randomized across participants.

## Results

The free parameters of both model versions were estimated for each participant based on their classification learning data using maximum likelihood. The resulting individual parameter values were used to predict the respective participant's classification behavior for the test stimuli.

The results clearly show that most participants ignored the within-category correlations. Specifically, participants tended to classify stimuli "T1-4" into category 1 and stimuli "T5-8" into category 0, in line with the predictions of the Euclidean similarity, see Figure 2. Mathematical modeling showed that, in the aggregate, the Euclidean similarity predicted participants' categorizations better than the Mahalanobis similarity, median log-likelihood across participants: Euclidean similarity model = -7.21, Mahalanobis similarity model = -21.63, random-choice model (predicting category probabilities of .50) = -16.64.

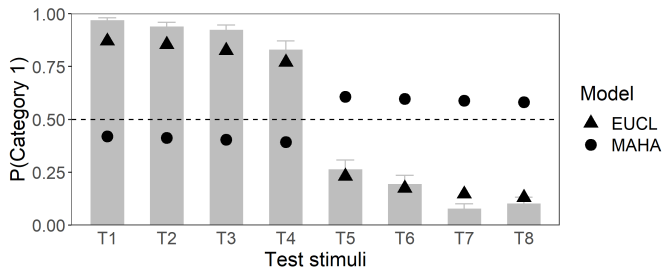


Figure 2: *Responses and predictions for the test stimuli.* Bars and whiskers show the mean and the standard deviation, respectively, across participants' mean category responses for a given test stimulus. The model predictions are aggregated over participants with the mean. EUCL = Euclidean similarity model; MAHA = Mahalanobis similarity model.

At the individual level, a participant was assigned to a model if the model's Akaike weight exceeded .67 for this participant. The results show most participants were best described by the Euclidean similarity ( $n = 33$  of 43) with strong evidence, see Figure 3. The remaining participants were described by the Mahalanobis similarity ( $n = 4$ ), the random-choice model ( $n = 2$ ), or unclassifiable ( $n = 4$ ).

## Conclusion

Our findings indicate that people do not include within-category feature correlations in their representation of similarity during categorization. Instead, people treat every object as stemming from a category with uncorrelated features, even if the true classification environment has strong within-category feature correlations.

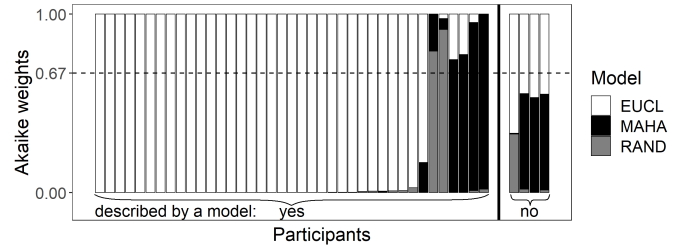


Figure 3: *Model evidence strengths for each participant.* For each participant, the Akaike weights of the different models are stacked upon each other and sum up to 1. EUCL = Euclidean similarity model; MAHA = Mahalanobis similarity model; RAND = random-choice model.

## Acknowledgments

This work was supported by the Swiss National Science Foundation [grant number P0BSP1\_195389 to F.I.S.].

## References

- Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition*, 30(3), 353–362. doi: 10.3758/bf03194936
- Ell, S. W., Smith, D. B., Peralta, G., & Hélie, S. (2017). The impact of category structure and training methodology on learning and generalizing within-category representations. *Attention, Perception, & Psychophysics*, 79(6), 1777–1794. doi: 10.3758/s13414-017-1345-2
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2), 125–157. doi: 10.1016/0010-0277(94)90065-5
- Lancaster, M. E., Shelhamer, R., & Homa, D. (2013). Category inference as a function of correlational structure, category discriminability, and number of available cues. *Memory & cognition*, 41(3), 339–353. doi: 10.3758/s13421-012-0271-8
- Mao, J., & Jain, A. K. (1996). A self-organizing network for hyperellipsoidal clustering (HEC). *Ieee transactions on neural networks*, 7(1), 16–29. doi: 10.1109/72.478389
- 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. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, 45(4), 279–290. doi: 10.3758/BF03204942
- Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10(2). doi: 10.5555/1577069.1577078