Beyond Distance and Density: Manifold Connectivity Shapes Similarity Judgments

Olivia A. Gatto[*1], **Andrew J. Steindl**[*2], **Smita Krishnaswamy**[2] **& Robert L. Goldstone**[1]
[1]Department of Psychological and Brain Sciences, Indiana University
1101 E. 10th Street Bloomington, IN 47405 USA
[2]Department of Computer Science, Yale University
51 Prospect Street New Haven, CT 06511
[*] Co-first authors contributed equally to this work.
ogatto@iu.edu, drew.steindl@yale.edu, smita.krishnaswamy@yale.edu, rgoldsto@iu.edu

Abstract

Learning new categories often alters similarity judgments: items from different categories are judged as less similar, and items within the same category as more similar. While category training may drive such changes, categories often have distributional structure without supervision. This raises the question of whether unsupervised exposure alone can reshape similarity judgments, and by what mechanism. We hypothesized that similarity judgments reflect a learned internal geometry of stimulus space. During unsupervised exposure, learners may infer the manifold structure underlying the distribution of examples, such that similarity depends on path-based relations on the manifold rather than inter-item distance or exemplar density alone. To test this, participants experienced stimulus distributions that differed only in whether they formed a connected or disconnected manifold. Despite equivalent physical similarity, participants judged items as less similar when separated by a topological gap. These results demonstrate that unsupervised exposure is sufficient to alter similarity judgments and reflect sensitivity to inferred geometric structure of stimulus space.

Keywords: unsupervised learning; categorization; manifold learning

Introduction

When people learn new categories, they subsequently tend to judge items belonging to different categories as less similar (Ashby et al., 2020; Bowman et al., 2022; De Leeuw et al., 2016; Goldstone et al., 2001; Livingston & Andrews, 2005; Pérez-Gay et al., 2017; Stevenage, 1998). In these learned categorical perception experiments, participants are given explicit category training and then they rate the similarity of within- and between-category pairs. By comparing pre-training or control group similarity ratings with post-training or learning group similarity ratings, researchers hope to ascertain whether and how the representational space of category examples changes as a result of category learning. Across numerous domains, participants judge stimuli crossing a category boundary to be less similar than those within a single category, consistent with the idea that category learning changes how stimulus dimensions are perceived and/or weighted.

This standard paradigm assumes that the driving force underlying representational shifts is explicit category training (e.g., labels, feedback, or rules), as opposed to the distributional structure of the examples. Nevertheless, category structure is often available in the absence of labels (Clapper & Bower, 1994). For instance, a trainee geologist might learn to differentiate between granite and diorite through instruction and feedback, but she could also notice certain combinations of rock features co-occurring among a set of unlabeled examples. One might imagine that the trainee could also "pick up" on the rock categories merely by studying the unlabeled examples and grouping them based on their features. This could lead to the trainee forming mental "clumps" of the rocks, and these clumps might correspond with actual categories as determined by geologists. More generally, one might imagine that individual, high-dimensional examples actually vary along a lower-dimensional manifold (e.g., a smooth, locally flat surface). Without labeling the examples, the task for the learner is to induce possible categories by learning the underlying manifold. This so-called unsupervised category learning takes place when learners notice common features between category members and group them accordingly, but it is as yet unclear whether unsupervised category learning can influence learners' judgments of similarity.

There are reasons to expect that distributional structure is important for impressions of similarity. Recent work shows that altering the internal distribution of category exemplars can directly modulate learned categorical perception effects on similarity ratings. For example, Silliman et al. (2020) demonstrated that introducing deviant exemplars into a category reduces within-category compression, plausibly by making the category appear less cohesive overall. Additionally, the authors report between-category compression as well, meaning that items crossing the category boundary were surprisingly rated as more similar after category training. Importantly, these effects occurred despite explicit category training and supervision, suggesting that representational change depends not only on explicit training, but also on the spatial organization of exemplars within the category distribution. Such findings motivate the hypothesis that learners are sensitive to distributional structure itself, and that this sensitivity may influence perceived similarity even in the absence of explicit training.

People seem to have strong prior beliefs about the distribution of category items, often assuming that category exemplars are organized around dense clusters of typical items with sparse tails of atypical items (e.g., that categories are unimodal and approximately normally distributed). When

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

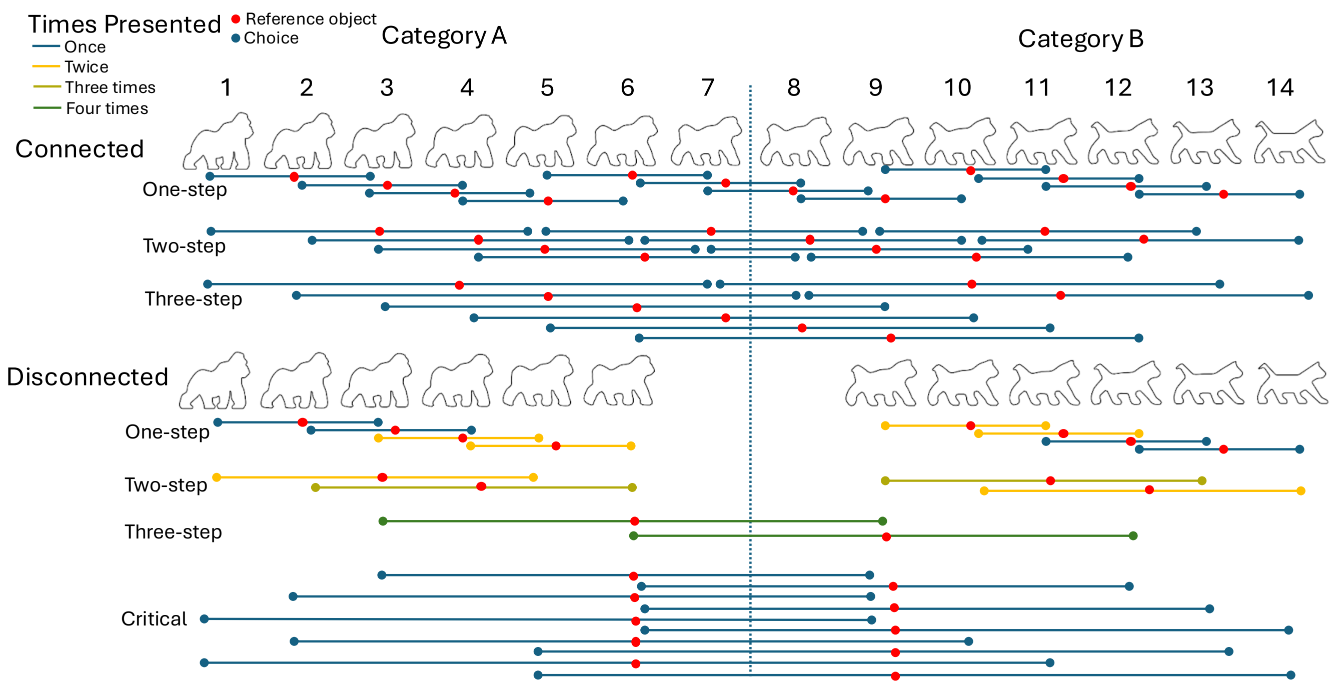

Figure 1: Example experimental stimuli. 8 morph sequences with 14 shapes each were created with unambiguous animal outlines as endpoints. In the Connected condition, forced choice similarity comparisons involved all 14 shapes, whereas in the Disconnected condition, all comparisons involving the center two shapes (7 and 8) were omitted. Comparisons involved one-step (e.g. 3 as a reference with 2 and 4 as choices), two-step (e.g. 3 as a reference with 1 and 5 as choices) and three-step (e.g. 4 as a reference with 1 and 7 as choices) differences. The critical items were the same for Connected and Disconnected conditions.

categories violate this prior assumption (such as when categories are defined by multimodal distributions of items), participants are slower to learn these categories (Flannagan et al., 1986; Fried & Holyoak, 1984). Relatedly, people make simple assumptions about category structure, such as there only being one linear category boundary, unless they are provided with strong evidence or hints about alternative, more complex manifold structures (Rogers et al., 2010). Even when knowledge about within-category clusters of items is incidental to categorization performance, people still seem to be sensitive to clusters such that items between clusters are made increasingly discriminable as a result of training, and items within clusters are made less discriminable (Gureckis & Goldstone, 2008). This is not merely explainable by participants weighing dimensions based on feedback on their categorization performance. If one dimension is merely being weighted higher as a result of category-training, then it is not clear why learners should show within-category cluster effects at all. Taken together, these results suggest that the distribution of category members in feature space might play a role in human impressions of similarity.

In the following experiment, we manipulate whether the stimuli are experienced as connected or disconnected manifolds while preserving pairwise Euclidean distances. We identify two competing predictions for how the manifold of category members might influence similarity assessments. The distance-density hypothesis predicts that when stimuli exist on a disconnected manifold, people will rate items from different clusters as more similar because, without intervening items, the clusters become naturally closer in representational space. Contrarily, when stimuli exist on a connected manifold, the presence of intervening items between clusters causes members from different clusters to appear less similar (Krumhansl, 1978). Alternatively, on a manifold/diffusion account, presenting stimuli on a disconnected manifold means that the path between clusters of items is broken, which leads to participants judging members from different clusters to be less similar. Meanwhile, when there is a connected manifold between clusters, the path between clusters through representational space leads members between clusters to appear more similar.

In the present research, we hypothesize that human similarity judgments are shaped by a learned internal geometry of the stimuli in feature space. During an unsupervised learning task, people infer a manifold structure, which they then use to generate similarity judgments. We argue that impressions of similarity depend on the path between examples on the manifold, over and above Euclidean distance and category item density. We demonstrate that the (Krumhansl, 1978) distance-density hypothesis does not provide accurate predictions of similarity judgments in our task. Then, we propose a diffusion model as a possible alternative, following tools such as PHATE that infer paths on the underlying data manifold.

PHATE Background

PHATE (Potential of Heat Diffusion for Affinity-based Transition Embedding; (Moon et al., 2019)) is a nonlinear dimensionality reduction method designed to preserve both local neighborhood structure and global manifold topology using diffusion geometry (Coifman & Lafon, 2006). Unlike commonly used nonlinear embedding methods such as t-SNE (van der Maaten & Hinton, 2008) and UMAP (McInnes et al., 2018), which prioritize local neighborhood structure and often distort global geometry, PHATE captures differences in global connectivity of the stimulus space even when Euclidean distances are held constant, aligning with our goal of modeling path-based similarity on the inferred manifold.

PHATE starts by constructing a k-nearest-neighbor graph over data points and then computes a diffusion operator from its edge affinities. Distances computed in this space approximate geodesic distances along the underlying data manifold (Coifman & Lafon, 2006). These distances are log transformed and embedded into low dimensions using multidimensional scaling.

Methods

Forty-seven participants were recruited from the Introductory Psychology Participant Pool at Indiana University in partial fulfillment of a course requirement. Thirty-one of the participants identified as female and 14 identified as male, and 2 preferred not to classify themselves by gender.

The stimuli consisted of 8 morph sequences created by starting with the outlines of two familiar animals as endpoints, and linearly interpolating between them using the Flubber library (https://github.com/veltman/flubber) to create sequences of 14 shapes. Care was taken in the choice of endpoints to avoid morph outlines that contained crossed lines. Examples of the morph sequences can be viewed at https://pc.cogs.indiana.edu/robsexperiments/tests-examples/morphcontours/index.html. The 8 sequences were defined by these endpoint pairs: coyote-bear, bear-sheep, sheep-coyote, llama-gorilla, gorilla-cat, sheep-otter, and camel-coyote, cat-llama.

The task consisted of participants making two-alternative forced-choice judgments, choosing which of two choice shapes is more similar to a reference shape. Figure 1 shows all of the comparison trials for each morph sequence. For each participant, 4 Connected and 4 Disconnected morph sequences were presented. Comparisons involving the center shapes 7 and 8 were included in the Connected but not Disconnected conditions. Comparisons included 30 trials on which the choices differed from the reference shape by one, two, or three positions on the morph sequence (one-step, two-step, and three-step comparisons, respectively). There were also 10 critical trials, identical for Connected and Disconnected condition, for which one choice belonged in the same shape category as the reference while the other choice was closer to the reference by zero, one, or two steps along the morph sequence. Call this *proximity advantage*.

Participants were instructed that on every trial they would see one shape at the top of the screen, and two shapes on the bottom-left and bottom-right. Their task was to decide whether they thought the left or right shape was more similar to the top shape. To make their judgment, they clicked with their mouse on the more similar shape, or pressed the "F" and "J" keys to respond left and right, respectively. Participants made 320 (30 X 8 morph-sequences) judgments in all, broken down into 8 blocks defined by a single morph sequence. For each morph sequence, participants were initially exposed to all of the shapes within the morph sequence for one second each so that they would get a feel for the range of possible shapes. Then, they received the 30 one-, two-, and three-step comparisons, randomly ordered and intermixed. Then, they received the 10 critical comparisons, randomly ordered. The two choices always involved one shape to the left of the reference shape in the morph sequence, and one shape to the right, but the left-right position of these two choices was randomized. The complete experiment can be experienced at: https://pc.cogs.indiana.edu/robsexperiments/manifoldsimilarity/index.html.

PHATE Modeling

To model the inferred geometry of the stimulus space, each morph sequence was embedded using PHATE. First we sampled 10,000 points uniformly from each stimuli curve. To ensure index alignment across consecutive shapes within a morph sequence, the point clouds were aligned by re-indexing the sampled points to minimize mean squared error between adjacent shapes. This alignment procedure preserves the intrinsic geometry of each shape, as it only changes the index set. The aligned point clouds were then flattened and used as input to PHATE. Unless otherwise specified, PHATE was applied with neighborhood size $k = 3$, and diffusion time t selected automatically according to the PHATE algorithm. Pairwise distances in the resulting embedding were used to derive predicted choice probabilities on critical trials via a softmax decision rule with $\beta = 5$. PHATE was treated as a deterministic model of the stimulus manifold rather than a fitted model of participant behavior.

Results

Behavioral Results

First, we examine the critical trials where the choices crossed into separate categories.

To model choice behavior at the trial level while accounting for repeated measures, we fit a binomial generalized linear mixed model with a random intercept for participant, predicting whether the same category option was chosen based on condition, proximity advantage, and their interaction. The proximity advantage exerted a large effect on choice, with the odds of choosing the same category option decreasing as proximity advantage decreased (1 step: OR = 0.80, 95% CI [0.72, 0.88]; 2 step: OR = 0.58, 95% CI [0.50, 0.66]). Critically, there was also a main effect of condition. Across proximity levels of the critical trials, participants in the Disconnected condition were more likely to choose the same category option than those in the Connected condition (OR = 1.27, 95% CI [1.16, 1.38]). The condition by proximity interaction was not significant, indicating that the effect of disconnection was additive rather than dependent on the magnitude of physical disadvantage. A repeated measures ANOVA on participant level choice proportions yielded the same qualitative pattern as the mixed-effects model, with significant main effects of proximity advantage and condition, and no interaction between them.

Zero proximity advantage trials. We fit a separate mixed-effects logistic model to the zero proximity advantage critical trials only (e.g., when the target shape was equidistant in morph space from its two reference shapes). Even when the within- and between-category options were matched in their morph sequence distance from the reference, participants were significantly more likely to choose the within-category option in the Disconnected condition than in the Connected condition (OR = 1.26, 95% CI [1.10, 1.46]) (Figure 2). The model showed a consistent shift favoring within-category choices when stimuli manifolds were disconnected (Connected: $P(within) = .56$; Disconnected: $P(within) = .61$). These results demonstrate that the connectivity of the experienced stimulus manifold biases similarity judgments when physical distance is held constant.

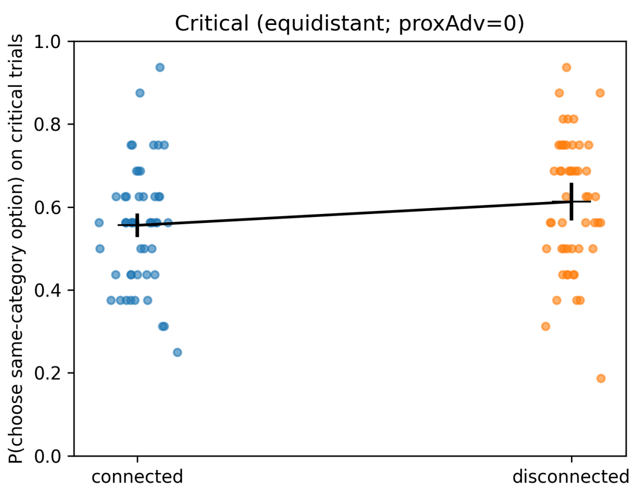

Figure 2: The predicted probabilities and 95% confidence intervals of equidistant critical trials showing a higher likelihood of choosing the same category option in the Disconnected than the Connected condition.

Robustness across morph sequences. To assess whether the condition effect was driven by individual morph sequences, we examined its consistency across sequences using a series of sequence-wise mixed-effects logistic regressions. For each morph sequence and each proximity advantage level (zero, one, and two), we fit a random intercept logistic model predicting within-category choice from condition. At the equidistant level (zero-step), five of eight sequences exhibited odds ratios greater than one, with the remaining sequences clustering near one. This consistency increased with disadvantage: seven of eight sequences showed odds ratios greater than one at both one-step and two-step levels. No single sequence exhibited an extreme effect capable of accounting for the overall condition difference. Instead, the pooled estimate across sequences fell within the range of individual sequence effects at each disadvantage level, indicating that the observed condition effect is distributed across morph sequences and not due to specific morph sequences.

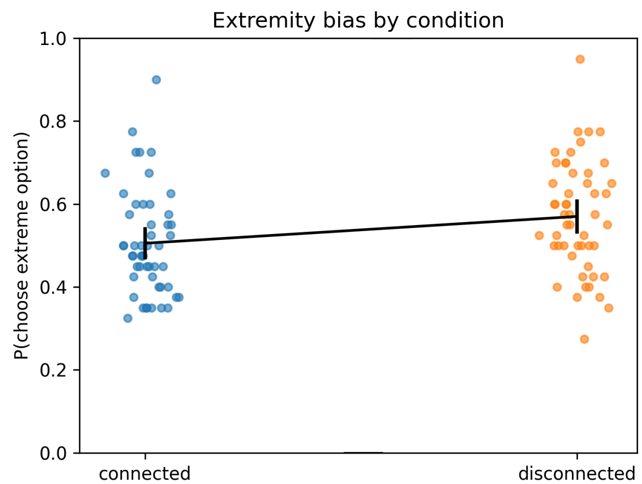

Figure 3: Extremity bias by condition. Subject level probabilities of choosing the extreme option (closer to 1 or 14), with condition means and 95% confidence intervals (black), excluding trials involving morph positions 7 and 8.

Extremity bias. Participants also exhibited a bias toward selecting more extreme exemplars in the Disconnected condition. Restricting analysis to trials excluding the central morph positions (7 and 8), a mixed-effects logistic regression revealed higher extremity choices in the Disconnected than the Connected condition (OR = 1.30, 95% CI [1.10, 1.53], p = .0017). Extremity choices were at chance in the Connected condition (p = .76) but above chance in the Disconnected condition (p = .0005), with a small effect size (ΔP = .065, Cohen's h = .13) where ΔP is the change in probability, Figure 3. This bias does not explain the primary connectivity effect, which remains on equidistant trials.

PHATE Modeling Results

We modeled similarity judgments using PHATE embeddings of each morph sequence. PHATE revealed a topological distinction between conditions. In the Disconnected condition,

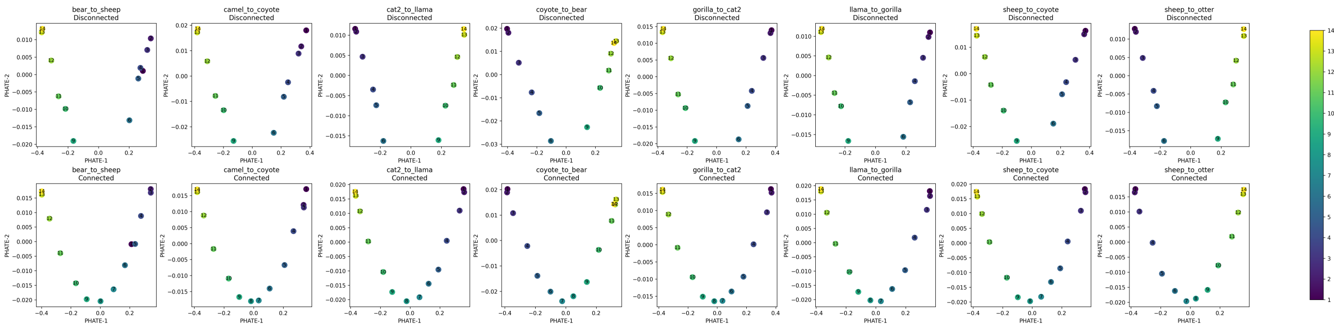

Figure 4: PHATE embeddings of morph sequences ($k = 3$, t = auto). Columns correspond to morph sequences. The top row shows the disconnected condition and the bottom row shows the connected condition.

morph sequences separated into two mostly distinct components, whereas in the Connected condition the same sequences formed a single continuous manifold (Figure 4). This difference in inferred connectivity arises despite Euclidean distances being experimentally controlled. Using pairwise distances in the PHATE embedding, we derived choice probabilities on critical trials via a softmax decision rule ($\beta = 5$). Across all proximity advantage levels, the model predicted a higher probability of choosing the option on the same category of the morph continuum as the reference in the Disconnected condition than in the Connected condition, mirroring the pattern observed in human behavior.

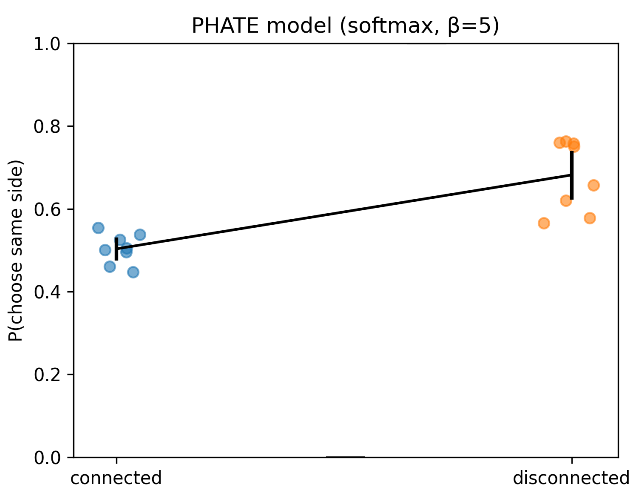

Figure 5: PHATE-inferred side choice probabilities on 0-step critical trials using a softmax decision rule ($\beta = 5$). Points indicate sequence-level probabilities aggregated across sequences; ticks denote mean differences between conditions. PHATE predicts higher same-side choice probability in the disconnected condition, paralleling human behavior.

PHATE Proximity Dependent Predictions For zero proximity advantage trials, PHATE predicted the strongest isolation advantage. The mean predicted probability of choosing the same category option was 0.503 ($SD = 0.044$) in the Connected condition versus 0.682 ($SD = 0.086$) in the Disconnected condition, resulting in $\Delta P = 0.179$ with ($F(1,30) = 55.11$, $p = 2.85 \times 10^{-8}$). The same pattern was observed consistently across morph sequences (Figure 5).

For one-step disadvantage critical trials, the predicted isolation advantage remained robust but was reduced in magnitude. The mean same category choice probability was 0.467 ($SD = 0.055$) in the Connected condition and 0.628 ($SD = 0.099$) in the Disconnected condition, corresponding to $\Delta P = 0.161$ ($F(1,30) = 32.22$, $p = 3{\times}10^{-6}$). Sequence-wise effects again showed consistent positive differences across morph sequences. For two-step disadvantage critical trials, the effect further decreased but remained reliable. PHATE predicted mean same category choice probabilities of 0.476 ($SD = 0.084$) for the Connected condition and 0.627 ($SD = 0.106$) for the Disconnected condition, yielding $\Delta P = 0.151$ ($F(1,30) = 19.96$, $p = 1.0 \times 10^{-4}$). Sequence-wise effects again showed consistent positive differences across morph sequences.

Robustness Finally, we evaluated the robustness of these predictions across neighborhood sizes ($k = 1 - 14$) and decision noise parameters ($\beta = 1 - 10$). Across a broad parameter range, PHATE-inferred choice probabilities exhibited the same qualitative pattern, with higher same-side choice probabilities in the Disconnected condition on critical trials. This effect reversed only at $k = \{7, 8\}$, corresponding to changes in the connectivity of the underlying k-NN graph. However, given the size of the stimulus set, with each side containing 5-8 category members, this reversal is to be expected because of the pigeonhole principle.

Discussion

We found evidence that participants in an unsupervised learning task are sensitive to the manifold structure underlying a set of examples, such that manifold structure impacted participants' impressions of similarity. When participants did not have experience with shapes that bridged between the two unambiguous animal endpoints, they were more likely to select items belonging to the same animal category as the referent shape as compared to the Connected condition, for which a traversable path through the two endpoint animals was available. This is in contrast to the distance-density hypothesis (Krumhansl, 1978), which posits that inter-item distance and overall density of category members are the predictive factors of similarity judgments, not the connectivity of an underlying manifold. In other words, while the distance-density hypothesis predicts that removing intervening items would make the resulting clusters seem more similar or confusable, we found evidence that a gap in the manifold makes objects spanning across the two clusters appear *less similar*. To our knowledge, this is the first demonstration of PHATE as a predictive model of human similarity choices in a cognitive task.

While human similarity judgments can be influenced by explicit category training (Ashby et al., 2020; Bowman et al., 2022; De Leeuw et al., 2016; Goldstone et al., 2001; Livingston & Andrews, 2005; Pérez-Gay et al., 2017; Stevenage, 1998), the present findings indicate that similarity judgments can also be systematically influenced by the structure of examples that learners are exposed to in the absence of labels or feedback. This has important implications for the study of learned categorical perception. Even without supervision, learners seem to infer category boundaries differently depending on the connectivity of the experienced stimuli manifold.

In addition to the primary connectivity effect, participants exhibited a stronger bias toward selecting more extreme category members in the Disconnected than Connected condition. Removing the bridging items in the stimuli manifold may have strengthened the endpoints as perceptual anchors, leading learners to rely more heavily on extreme exemplars when the continuity of the space was disrupted. Importantly, this extremity bias does not account for the full effect of manifold connectivity. On critical trials, where one item was more extreme than the other in both the Connected and Disconnected conditions, participants still showed a stronger bias favoring within-cluster choices in the Disconnected than Connected condition. This indicates that manifold connectivity exerts an influence on similarity judgments beyond any general preference for extreme exemplars. Nevertheless, an important direction for future research will be to examine whether the effect of manifold connectivity generalizes to unfamiliar stimulus domains, where anchoring to known category endpoints is minimized. Future experiments are needed to disentangle whether a general preference for clearer, more extreme shapes partly drives the effects of manifold connectivity, or if perceptual space is directly warped as a result of the space of experienced examples, such that shapes in the middle of the space are warped towards the endpoints more so in the Disconnected than Connected case.

The PHATE modeling results may provide insights into the mechanisms driving the observed effects of manifold connectivity. A broad parameter sweep revealed that the qualitative pattern of model predictions was robust across neighborhood sizes and β values, indicating that the predicted same-cluster bias in the Disconnected condition is governed primarily by inferred manifold connectivity.

The $k = 3$ k-NN graph used in our primary analyses represents a rough but standard approximation of the underlying stimulus structure. An alternative graph representation of each morph sequence would consist of a simple chain, in which each node in the middle of the chain would have two neighbors, and each endpoint would have one. This representation could be extended such that the Disconnected condition would have no edges bridging the central gap, making the manifolds strictly separated and yielding undefined or infinite distances between inter-category members. In the current graph representation, $k = 3$ means that clusters in the Disconnected condition share fewer edges while still being connected. Because PHATE relies on a k-nearest-neighbor graph and diffusion, separation between disconnected components is softened, resulting in graded similarity predictions instead of deterministic ones.

We did not attempt to explicitly model the observed extremity bias, as it was secondary to our primary hypothesis. However, a simple chain graph would naturally concentrate diffusion mass at endpoints due to reduced degrees of freedom, suggesting a potential mechanism by which manifold geometry could also give rise to extremity preferences. The simple chain graph would allow future modeling efforts that account for both the extremity bias and manifold connectivity effects. In contrast, with a uniform k-nearest-neighbor graph, such boundary effects are attenuated, making PHATE less sensitive to extreme choice tendencies under the present parameterization.

Borrowing manifold learning techniques from diffusion geometry, we examined whether the underlying structure of encountered examples influences one's subsequent judgments of similarity. The manipulation of removing two intervening items from a randomized sequence of stimuli was sufficient to induce a within-cluster bias, even holding inter-item distance and category density constant. Additionally, using PHATE, we modeled these behavioral results by taking the individual items as lying on a lower-dimensional manifold. We put forth manifold learning as a viable candidate for partly explaining the influence of distributional properties on human judgments of similarity.

Acknowledgments

An AI language model (ChatGPT, OpenAI) was used to assist with general writing and coding assistance.

References

Ashby, S. R., Bowman, C. R., & Zeithamova, D. (2020). Perceived similarity ratings predict generalization success after traditional category learning and a new paired-associate learning task. *Psychonomic Bulletin & Review*, *27*(4), 791–800. https://doi.org/10.3758/s13423-020-01754-3

Bowman, C. R., Ashby, S. R., & Zeithamova, D. (2022). Age effects on category learning, categorical perception, and generalization [Epub 2021 Nov 11]. *Memory*, *30*(3), 230–247. https://doi.org/10.1080/09658211.2021.2003818

Clapper, J. P., & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 443–460. https://doi.org/10.1037/0278-7393.20.2.443

Coifman, R. R., & Lafon, S. (2006). Diffusion maps [Special Issue: Diffusion Maps and Wavelets]. *Applied and Computational Harmonic Analysis*, *21*(1), 5–30. https://doi.org/https://doi.org/10.1016/j.acha.2006.04.006

De Leeuw, J. R., Andrews, J. K., Livingston, K. R., & Chin, B. M. (2016). The effects of categorization on perceptual judgment are robust across different assessment tasks. *Collabra*, *2*(1), 9. https://doi.org/10.1525/collabra.32

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 241.

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.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43. https://doi.org/10.1016/S0010-0277(00)00099-8

Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. *Proceedings of the 30th annual conference of the cognitive science society*, *10*.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *85*(5), 445–463. https://doi.org/10.1037/0033-295X.85.5.445

Livingston, K. R., & Andrews, J. K. (2005). Evidence for an age-independent process in category learning. *Developmental Science*, *8*(4), 319–325. https://doi.org/10.1111/j.1467-7687.2005.00419.x

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction [cite arxiv:1802.03426Comment: Reference implementation available at http://github.com/lmcinnes/umap]. http://arxiv.org/abs/1802.03426

Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., Yim, K., Elzen, A. v. d., Hirn, M. J., Coifman, R. R., Ivanova, N. B., Wolf, G., & Krishnaswamy, S. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, *37*(12), 1482–1492. https://doi.org/10.1038/s41587-019-0336-3

Pérez-Gay, F., Thériault, C., Gregory, M., Sabri, H., Rivas, D., & Harnad, S. (2017). How and why does category learning cause categorical perception? *International Journal of Comparative Psychology*, *30*. https://doi.org/10.46867/ijcp.2017.30.01.01

Rogers, T., Kalish, C., Harrison, J., Zhu, J., & Gibson, B. (2010). Humans learn using manifolds, reluctantly. *Advances in neural information processing systems*, *23*.

Silliman, D., Snoddy, S., Wetzel, M., & Kurtz, K. J. (2020). Costly exceptions: Deviant exemplars reduce category compression. In S. Denison, M. Mack, Y. X. 0023, & B. C. Armstrong (Eds.), *Proceedings of the 42th annual meeting of the cognitive science society - developing a mind: Learning in humans, animals, and machines, cogsci 2020, virtual, july 29 - august 1, 2020*. cognitivesciencesociety.org. https://cogsci.mindmodeling.org/2020/papers/0458/index.html

Stevenage, S. V. (1998). Which twin are you? a demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology*, *89*(1), 39–57. https://doi.org/10.1111/j.2044-8295.1998.tb02672.x

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, *9*(86), 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html