

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Uncovering the Metricity of Representational Spaces in the Brain: Evidence from Colors and Letters

Permalink

<https://escholarship.org/uc/item/6r11x1md>

Journal

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

ISSN

1069-7977

Authors

Caglar, Leyla
Hanson, Catherine
Hanson, Stephen Jose

Publication Date

2021

Peer reviewed

Uncovering the Metricity of Representational Spaces in the Brain: Evidence from Colors and Letters

Leyla Roxsan Caglar (lcaglar@andrew.cmu.edu)

Department of Psychology, Carnegie Mellon University, 4909 Frew Street
Pittsburgh, PA 15213 USA

Department of Psychology, University of Coimbra, R. Inácio Duarte 65
Coimbra, 3000-481 Portugal

Catherine Hanson (cat@rubic.rutgers.edu)

Department of Psychology, Rutgers University Brain Imaging Center (RUBIC), Rutgers University, 197 University Avenue
Newark, NJ 07102 USA

Stephen José Hanson (jose@rubic.rutgers.edu)

Department of Psychology, Rutgers University Brain Imaging Center (RUBIC), Rutgers University, 197 University Avenue
Newark, NJ 07102 USA

Abstract

An ongoing debate about the structure of conceptual space is based on two competing mathematical theories of similarity that make distinct predictions about the structure of mental representations and how to model the representational space they are stored in. These are known as metric (Shepard, 1962) and ultrametric (Tversky, 1977) theories, modeled by multidimensional scaling and additive trees respectively. Turning to the brain to resolve this conflict, we propose a computational framework to assess behavioral and neural data's underlying structure and investigate whether the behaviorally known spaces for colors (metric) and letters (ultrametric) can be reproduced from neural data. Our results show that the metric color wheel can be reproduced from brain area V4, but that neural activations of the letters from extrastriate cortex (V2-V5) are also metric instead of being ultrametric. Finally, we discuss three possibilities for the brain's similarity structure, including a potential metric bias.

Keywords: additive tree; concepts, MDS; metricity; representational spaces; similarity structure, ultrametric

Introduction

The brain can represent a vast amount of information about objects, concepts, and categories at different levels of abstraction while also preserving their individuality and relationships to one another. Understanding how the brain solves such fundamental problems of representation is still subject to avid debate. Our approach in tackling these queries is to characterize different properties of the conceptual (or psychological) space, thought to store all representations of object, concept, and category knowledge.

Central to this endeavor is the notion of similarity – a relation between two perceptual or conceptual objects – which fundamentally underlies our understanding of concepts and categories and their mental representations (Edelman, 1998; Hahn & Chater, 1997; Shepard, 1980, Tversky, 1977). Perceived similarity has classically been measured using behavioral tasks, such as similarity judgments, object sorting, substitution errors, or

identification/discrimination errors. This idea of measuring an item's internal representation via its relation/similarity to other items is based on a second-order isomorphism of internal representations (Shepard & Chipman, 1970). A second-order isomorphism - also used in representational similarity analysis (RSA; Kriegeskorte et al., 2008) - assumes no direct correspondence of properties between an object's external and internal relationship, allowing for measurements of similarity without prior knowledge about the physical stimuli's structure or properties.

Mathematical Models of Similarity

In the past century, two prominent mathematical theories have made distinct predictions about the organization of mental representations and how to model the psychological (or conceptual) space they are stored in.

Metric theories (Shepard, 1962; Thurstone, 1927) propose that concepts are represented as points in a continuous metric space. The similarity (or dissimilarity) between concepts is argued to be a function of the distance between all other concepts, with more similar concepts being represented by closer points in the space. Proposed as a Universal Law of Generalization, Shepard (1987) argues that perceived similarities are inversely related to psychological distances via an exponential decay function of form $s(A,B)=e^{-d(A,B)}$, where A and B are two percepts (e.g. objects or concepts). This metric psychological space can be modeled and visualized using a technique called non-metric multidimensional scaling (MDS) (Kruskal & Wish, 1978). The goal of MDS is to reconstruct a k-dimensional geometric space that preserves the distances of perceived (dis)similarities.

Ultrametric theories (Tversky, 1977; Tversky & Gati, 1978), on the other hand, propose that similarity is a function of both common and distinctive features. Here, perceived similarities are a result of a feature matching process of common and distinctive features that are differentially weighted, leading to a concept with more

common and less distinctive features to increase in similarity. Tversky's contrast model expresses this as a linear combination, where the similarity of stimuli A and B is $s(A, B) = \alpha g(A \cap B) - \beta g(A - B) - \gamma g(B - A)$, with the constants α , β , and γ being subject to contextual change. Tversky's contrast model can visually be represented in a variety of different tree structures (i.e., hierarchical clustering schemes, and ultrametric or additive trees). Yet, Sattath and Tversky (1977) directly extended the contrast model through the construction of additive feature trees in a method called Addtree, which is an unrooted binary additive tree, where concepts (and their shared features) are represented as nodes in a connected graph and relations between nodes reflect the observed proximities.

Metric and ultrametric theories (also referred to as spatial and tree models respectively) describe different cognitive processes involved in determining two objects' similarity. Detecting whether a dataset is better fit by a metric or ultrametric model can consequently be informative about the cognitive processes that underly object perception, similarity, categorization, and representation.

Behavioral Evidence

Ample behavioral evidence supports both metric and ultrametric theories as a characterization of psychological space (Pruzansky, Tversky, & Carroll, 1982; Shepard, 1980) and as a tool to reveal the underlying structure and dimensions of various representational spaces. Shepard (1964), for example, showed that a 2D MDS solution of Ekman's (1954) color data recovers the well-known color wheel (Newton, 1704). Meanwhile, researchers were also testing the ultrametric theory and the fit of tree models (Tversky & Gati, 1982). Sattath and Tversky (1977) compared the MDS and Addtree solutions of visual similarity judgments of Swedish letters, originally collected by Kuennapas and Janson (1969). While the MDS solution reveals one possible dimension of round vs. straight letters, the other dimensions remain ambiguous. The Addtree solution however reveals very distinct and easily interpretable clusters (e.g., vertical, arched, etc.).

To date, the investigation and comparison of metric and ultrametric theories has exclusively been based on behavioral data and researchers were not able to resolve the conflict, concluding that both the task and the stimulus structure determine what the structure of the representational space will be (Pruzansky, Tversky, & Carroll, 1982). This might be because behavioral similarity data is biased and subjective, subject to change based on the context and stimulus sample, and limited in dimensionality. Furthermore, behavioral similarity data might be a more indirect way of measuring representations, as it is the output of a sequence of decision processes that can undergo a variety of cognitive transformations before being measured as behavioral output. Turning to the brain could provide new constraints in differentiating between these theories, offer higher dimensionality and degrees of freedom, and constitute a more direct way of measuring representations

and their representational spaces before undergoing some of the cognitive processes.

Consequently, our goal was to experimentally investigate whether known similarity representations underlying psychological space can be reproduced from neural data. Using functional magnetic resonance imaging (fMRI), we investigated whether the known psychological spaces of colors (metric) and letters (ultrametric) can be modeled from neural activations, predicting that their representational spaces will be congruent.

Detecting Similarity Structures

To evaluate the representational structure of conceptual spaces, we propose a computational framework that assesses the performance of the metric and ultrametric theories' modeling tools (i.e., MDS and Addtree). In addition to providing reliable measures of whether a given data is better represented in a metric or an ultrametric space, this framework also needs to be applicable to both behavioral and neural data.

The proposed similarity structure analysis (SSA) consists of three different measures and is an extension of the comparative methods proposed by Pruzansky, Tversky, and Carroll (1982). The first measure is a visual examination and interpretability of the data in the metric space using nonmetric MDS (Kruskal & Wish, 1978; Kruskal, 1964) and in the ultrametric space using least square fits of Addtree (Sattath & Tversky, 1977). Despite the lack of explicit feature dimensions, these visualizations can be examined for shared features to identify the data's underlying dimensions. The second measure evaluates the models' goodness of fit. Model fits are determined based on the r^2 -value based on a linear regression between observed proximities and model distances as well as by computing and comparing the Akaike Information Criterion (AIC) of each model. The third measure examines the data's skewness based on standardized and centered frequency distributions. This is based on Sattath & Tversky's (1977) Skewness Theorem, which shows that left-skewed data is better modeled by ultrametric spaces and right-skewed data is better modeled by metric spaces.

To demonstrate the SSA's validity we used two different data types with a known structure: real-world metric distances between US cities and taxonomic data of plants (Oksanen et al., 2019). The SSA shows an accurate detection of the underlying similarity structure in both cases (Table 1). Visually, MDS scales the US city distances into their correct positions on a map of the US, while Addtree separates the cities into east and west coast locations. While Addtree shows a relatively good r^2 fit as well, only MDS is able to uncover the underlying latent structure of the US map, highlighting the importance of considering all three tests in conjunction. For the taxonomic data of plants, on the other hand, Addtree outperforms MDS and provides a perfect fit to the data ($r^2=1$) and groups them by their ecological constraints.

Second, we tested the SSA’s performance on behavioral similarity judgments with an unknown structure, using the well-known case of colors and letters (Table 1). As expected, the SSA reveals a better metric fit for the colors and produces the well-known color wheel (Ekman, 1954; Newton, 1704; Shepard, 1964), while the letters are better fit in an ultrametric space with Addtree separating the letters into clusters based on their visual shapes (Kuennapas & Janson, 1969; Sattath & Tversky, 1977).

Behavioral Experiments

Participants & Stimuli

Participants Fifteen participants (mean age = 26.3, SD; female = 6) were recruited to perform a behavioral version of the Ekman (1954) color experiment. All participants were prescreened for normal color vision using the Ishihara colorblindness test (Ishihara, 1917). Six participants (mean age = 25.1, female = 2) were recruited to perform a behavioral version of the Kuennapas and Janson (1969) letter experiment. Participants were prescreened for dyslexia and provided behavioral similarity ratings while performing an fMRI experiment. All participants gave informed consent in accordance with the Rutgers University Institutional Review Board and either volunteered or received course credit for their participation.

Color Stimuli Fourteen square color stimuli (31x31 pixels) were created based on the original wavelengths (434nm, 445nm, 465nm, 472nm, 490nm, 504nm, 537nm, 555nm, 584nm, 600nm, 610nm, 628nm, 651nm, 674nm) used in the Ekman (1954) study. To digitalize the color stimuli, we converted the wavelength values into the tristimulus color space CIE XYZ to closely resemble the perception of colors by the human eye.

Letter Stimuli The letter stimuli consisted of the same set of lowercase Latin alphabet letters as in the Kuennapas and Janson (1969) study, with the exception of three Swedish letters (å, ä, ö) that do not exist in the English alphabet and “w”, which is not part of the Swedish alphabet.

Experimental Procedure

Table 1: Results of the SSA for US cities, plants, and the original color (Ekman, 1954) and letter (Kuennapas & Janson, 1969) data, as well as their behavioral replications.

Data	MDS r^2	Addtree r^2	Data Skew	MDS AIC	Addtree AIC
US Cities	1	0.97	0.27	260	567
Plants	0.82	1	0.9	4563	-27509
Colors	0.94	0.86	-1.16	-247	-171
Letters	0.55	0.77	1.36	-433	-631
<u>Beh. Data</u>					
Colors	0.96	0.84	0.7	27	92
Letters	0.65	0.75	-0.75	949	845

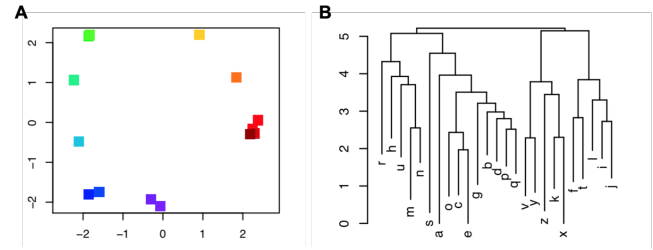


Figure 1: Grouped averaged behavioral results for the color MDS (A) and letter Addtree (B) solutions.

To get a sense of the distribution of colors, participants were first familiarized with all color stimuli and performed a warm-up task identical to the experimental task using four shades of grey tones. For the main experiment, participants were instructed to rate the similarity between all possible color pairs presented in a random order. A fixation cross was coupled with an ISI jitter (1-2s), followed by a 6s presentation of color pairs (first 2s by itself, then 4s with the similarity scale) and a similarity scale that appeared after the first two seconds of the color pair presentation (2s; 1-5 with 1 = no similarity; 5 = identical). The order of pairs was randomized for each participant. All experimental tasks were presented using the experimental presentation software PsychoPy2.

The letter task was identical to the design of the color task, except that participants were instructed to rate the *visual* similarity between all possible letter pairs on a Likert scale of 1-10 (1 = no similarity; 10 = identical) in accordance with the original study design.

Analysis

Each participant’s similarity judgments were converted into a distance (or dissimilarity) matrix to serve as input to the SSA. A group average was computed by taking the mean over participants’ distance matrices. Additionally, Kendall’s Tau was used to facilitate a direct comparison between our replications and the original study data.

Results

Color Experiment Both within single subjects and the group mean, all participants’ MDS configurations exhibited a concentric structure and correct ordering of the color wheel (Figure 1A). The results of the SSA shows that the MDS’ metric space is a better fit to the data. A comparison of the group dissimilarity matrix based on the behavioral data and the dissimilarity matrix of the original Ekman data using Kendall’s Tau rank correlation coefficient further reveals that they are highly and significantly similar ($r_t = 0.75, p < .0001$).

Letter Experiment Participants’ Addtree representations, both on the individual and on the averaged group level, exhibited a clustering congruent with the different shape categories of letters (Figure 1B), replicating the results of the original letter data. The SSA results confirm that the ultrametric space is a better fit (Table 1) and significantly

similar to the original Kuennapas and Janson data (1969) ($r = 0.68$, $p < .0001$).

fMRI Experiments

The results of the behavioral experiments show a successful replication of the color and letter experiments conducted half a century prior. Next, we adapted the same experiments for fMRI data collection to examine whether the brain represents colors and letters in the same similarity structure.

Participants & Stimuli

Four participants were recruited for the color study and prescreened for possible colorblindness using the Ishihara color test. Another four participants (mean age = 37.5; female = 3) with no history of dyslexia or reading impairments participated in the letter study.

Experimental Procedure

Colors To assure that participants were familiar and consistent enough with using the similarity scale, they received 2-3 training runs of the task prior to performing it in the fMRI scanner. The task consisted of two parts presenting the color stimuli with different experimental designs. In the first part, participants saw each color stimulus individually in a block design. Each trial consisted of first seeing either a cross or a hash (coupled with an ISI of 3-4s) and then a color stimulus (6s). As an attention check, participants were instructed at each trial to indicate whether they had seen a cross or a hash by button press. The order of both the colors and the cross/hash presentations was randomized and repeated over two separate runs. The second part was identical to the behavioral experimental design and consisted of rating the similarity between all possible color pairs.

Letters The letter task consisted of two parts, in which participants were first asked to rate the similarity of randomized letter pairs (4s of simultaneously presented pairs and rating scale). Then, after a fixation cross combined with a jitter (0.5s), one of the letters in the pair was presented in the center of the screen and participants had to indicate the previous position of the stimulus in the letter pair (4s). There were 13 individual presentations per letter and the side of presentation in each letter pair was counterbalanced.

fMRI Data Acquisition All neuroimaging data was collected at the Rutgers University Brain Imaging Center with a Siemens Trio 3T scanner and a 32-channel head coil. The experimental tasks were presented using PsychoPy3 and displayed on a screen visible during fMRI scanning. Participants' responses were collected via presses on a fMRI compatible button box. We collected anatomical MRI data (T1-weighted 176 1mm slices) as well as functional data, based on 35 contiguous slices (1mm³ voxel resolution) set parallel to the AC-PC plane during the experimental task, and using a multiband sequence (TR = 1s). Preprocessing

(spatial normalization to standard MNI space) and analysis of fMRI data was conducted in FSL and R.

Analysis

Anatomically defined region of interest (ROIs) masks were created using the Juelich Historical atlas and thresholded at 60% probability. Based on their association with color representation (Bohon et al., 2016) and single letter representation (Flower et al., 2004), we chose the target ROIs of V4 and extrastriate cortex (V2-V5) respectively. The auditory cortex served as a control ROI for both studies. We extracted the time series activations for each ROI and averaged over the third 2s of presentation associated with each condition's individual stimulus presentation. This resulted in a single activation vector per stimulus, capturing the spatial activations over all voxels of the ROI. Since all activation vectors were highly correlated with each other, we applied a novel two-step de-noising method to enhance the signal to noise ratio described (see Hanson, Caglar, & Hanson (2020) for details). In the first step, we created an independent noise sample using the neural activity associated with the fixation time points immediately preceding the individual stimulus presentations. Then, we used this independent noise sample and applied it as an orthogonal voxel filter to the stimulus voxel matrix. This procedure successfully de-correlated the original stimuli from their background by more than 80%, allowing for accurate stimulus decoding. We then computed Euclidean distance matrices to input to the SSA. To compare the representational shapes and organizations we used entanglement values, cophenetic correlation coefficients, and Procrustes analysis.

Results

Colors For all four subjects, individual MDS solutions based on the neural activations of V4 revealed the expected color wheel with continuous and correct ordering of hues (Figure 2A). The model fits of the SSA further reveal that the metric MDS space is a better fit to the data in all cases (Table 2). This was also confirmed by an average Kendall's Tau coefficient of 0.57 when compared to the original Ekman data. Importantly however, the same color wheel-like structure could not be found for the fixation activations in the auditory cortex control ROI (Figure 2A).

Letters Contrary to the behavioral data, the representational space derived from activations in extrastriate cortex do not show the expected clustering based on visual shape. Instead, the letters are ordered continuously and according to the alphabet in each participant's MDS and Addtree spaces (Figure 2B). Interestingly, the Addtree clustering of the space is in line with the phrase structure of the alphabet song, a common mnemonic used to learn the alphabet at a young age (see Hanson, Caglar, & Hanson, 2020 for details and discussion). Also diverging from the behavioral data, the neural representation of letters is better fit in a metric space than in an ultrametric space (Table 2). Unsurprisingly,

Table 2: SSA results for the color and letter fMRI studies.

Subjects	MDS r^2	Addtree r^2	Data Skew	MDS AIC	Addtree AIC
Colors1	0.89	0.77	1.01	2394	2458
Colors2	0.94	0.84	0.41	2031	2115
Colors3	0.9	0.82	0.63	1471	1525
Colors4	0.88	0.79	0.8	1372	1423
Letters1	0.90	0.75	1.23	4781	5048
Letters2	0.89	0.77	1.12	4724	4935
Letters3	0.88	0.70	1.85	4693	4957
Letters4	0.85	0.68	1.89	7475	7701

the average Kendall’s Tau also reflects this lack of congruency between the behavioral and the neural data ($r_t = 0.11, p = 0.06$).

Higher Dimensional Tests Even though the letters’ 2D MDS solution looks circular and similar in shape to the MDS configurations of the colors (Figure 2B), plotting the first three dimensions of a 6D MDS solution clearly reveals a consistent shape across participants that is not circular, but forms a figure eight (Figure 2B). This choice of 6D is based on the number of clusters produced by Addtree. The geometry and curvature of the helical folding follows the phrase structure of the alphabet song as apparent by the color coding of the clusters (Figure 2B). This is in contrast to the color data, which, plotted as a 3D projection of a 6D MDS solution, does not produce the same helical shape but maintains the data’s more circular shape even in higher dimensions (Figure 2A).

To further exclude the possibility of an artifact, we ran a simulation by sampling from a Gaussian random variable. We matched the average mean and standard deviation of the subjects’ denoised fMRI letter data, computed a distance matrix, and scaled it into 6D using MDS. Resampling over six separate simulations, the resultant plotting of a 3D projection of the 6D space shows that neither the sequential ordering of the letters, nor the helical structure is present (Figure 2C).

Control ROIs Although not as clean as in extrastriate cortex, the letter stimulus activations in auditory cortex also show an alphabetical organization in Addtree, probably because of the strong association with the alphabet song. Plotted as a 2D MDS configuration (Figure 2C), the letters at first look out of order. However, closer investigation reveals that it is a flattened spiral structure, which causes the letters to be intertwined in their order when projected on top of each other in 2D. A 6D MDS solution plotted in 3D enables clear identification of this shape (Figure 2C). Importantly however, neither the alphabetical ordering, nor the geometric structure can be found in the fixation activations of the auditory cortex (Figure 2A & B).

Contrary to the extrastriate results, in the auditory cortex the ultrametric space is a slightly better fit to the data ($r^2 = 0.45$; AIC = 4025.0; skew = 0.923) than the metric space ($r^2 = 0.51$; AIC = 3987.75). We suspect that this is because the 2D MDS solution necessarily conflates the distance between letters, while Addtree is still able to draw up clusters.

The overlap between representations from extrastriate cortex and auditory cortex is characterized by low entanglement coefficients both across individual subjects ($E_{\text{mean}} = 0.23$) and on the group average ($E = 0.33$). A cophenetic correlation matrix further reveals stronger average correlations in clustering within the extrastriate cortex (CPCC = 0.67) and the auditory cortex (CPCC = 0.445), than between the two ROIs (CPCC = 0.376).

Lastly, we performed Procrustes analyses of the MDS configurations’ shapes both across subjects within the same ROI and within subjects across ROIs. All analyses were run over 999 permutations to help determine significance in the shape’s correlations. Within ROIs, all subjects were compared to each other, revealing that their representations in extrastriate cortex were 0.973 correlated ($\text{rmse}_{\text{mean}} = 0.054$; $p_{\text{mean}} < 0.001$). Even though some of the subject’s representational shapes in auditory cortex did show concordance, the correlations were significantly lower at ~ 0.5 ($\text{rmse}_{\text{mean}} = 0.691$; $p_{\text{mean}} = 0.075$). Contrary to the within subjects’ analysis of each ROI and with an average

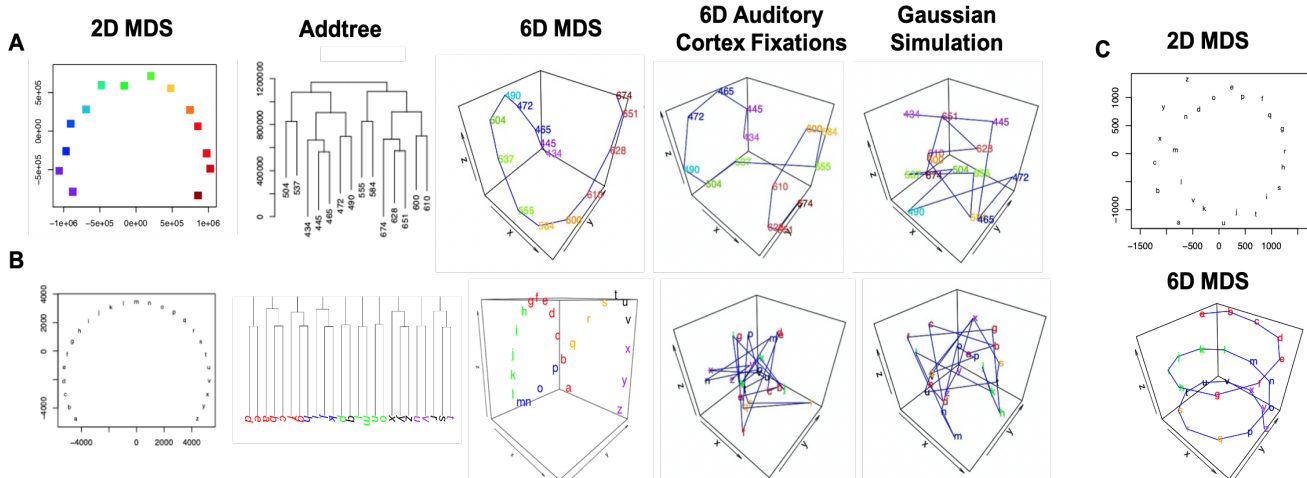


Figure 2: MDS and Addtree representation for a single subject of the color (A) and letter (B) study. 2D and 6D (plotted in 3D) MDS solutions of the letter data in auditory cortex showing a helical shape.

of only 0.33 correlation, none of the subjects except for one showed a significant correlation between the shape of their representation in extrastriate cortex and auditory cortex ($\text{rmse}_{\text{mean}} = 0.852$; $p_{\text{mean}} = 0.455$).

Discussion

To resolve the longstanding conflict between the metric and ultrametric theories, we examined whether known psychological spaces from behavioral data can be reproduced from neural fMRI data. We accomplished this by developing a computational framework that can successfully detect the underlying metricity of representational spaces across brain and behavior.

Using the behavioral case of metric colors (Ekman, 1954; Shepard, 1962) and ultrametric letters (Kuennapas & Janson, 1969; Sattath & Tversky, 1977), our behavioral studies show half a century later that their representational spaces and similarity structures can successfully be replicated in a computerized version.

Using the same color stimuli, we further show in the color fMRI study that the known metric similarity structure of colors can also be reproduced from neural data in V4. Confirming our hypothesis, the behavioral and neural representational spaces of color are both better fit in a metric space and show high congruency. Surprisingly, however, extrastriate cortex neural activations of the behaviorally ultrametric case of letters exhibit an alphabetical ordering with a better metric fit. This shows a strong incongruency of behavioral and neural metricity for the letters, leading both neural datasets to exhibit a metric favoring.

We first address potential alternative causes for the neural letter data's metricity, before providing various interpretations and implications of these results. An initial concern is related to the difference in task demands in the behavioral and neural experiments. While the behavioral task involved participants providing similarity judgments, neither of the fMRI tasks modeled activations from similarity comparisons. Instead, the activations were obtained either during passive viewing of individual color stimuli or while viewing individual letter stimuli and indicating their previous position. Despite the difference in behavioral and neural tasks, the color condition nevertheless shows equivalent representational spaces. While this does not exclude potential effects that different experimental tasks might have on the shape and organization of the representational space, it does provide evidence that the metricity results of the neural letter task might not be due to such task difference. In order to assure that the better metric fit and alphabetical ordering of the neural letter data is not caused by task modulation, follow up experiments should include an identical paradigm to the neural color experiment, as well as other experimental designs such as using rotated or inverse letters to enhance the saliency of the letters' visual shapes.

A second concern is related to the appropriateness of using metric and ultrametric spaces. It is important to point out that an ultrametric space is a special case of a metric space that is more restrictive by strengthening the triangle inequality through either the ultrametric or additive inequality. Consequently, a good ultrametric fit would always imply a metric space, whereas a good metric fit does not necessarily entail an ultrametric space. While this might be a crucial detail for interpreting results of other datasets, we do not believe that it holds explanatory power in the case of the neural letter results, which was hypothesized to be better fit in an ultrametric space but instead is better fit in the less restrictive metric space.

Relatedly, the appropriateness of both the metric and ultrametric models has been questioned and contented in the past - amongst others, for assuming symmetry (Tversky, 1977) or for not adequately distinguishing between relational and feature similarity (Goldstone, Medin, & Gentner, 1991). These critiques are most pertinent with different task demands and the relation between different category and exemplar sets (Medin, Goldstone, & Gentner, 1993; Roth & Shoben, 1983) and thus will play an important role in more complex experimental designs than those used in the here presented studies. Additionally, it should not affect the consistency of representational spaces between the behavioral and neural data, which was our main focus of comparison.

We offer three possible interpretations of the neural results. A first possible interpretation could be that the brain generally represents concepts in a metric similarity structure. The ultrametric spaces observed in behavior data could be a result of behavioral data being a more indirect measure of representations and undergoing transformations by other task-related cognitive processes.

A second interpretation relates to claims made by Pruzansky and colleagues (1982), who proposed a perceptual-conceptual dichotomy. Based on a near equal numbers of datasets with perceptual (e.g., forms, colors, and sounds) and conceptual content (e.g., animals, tools, sports), the authors claimed that overall, perceptual data tends to be better represented in a metric space, while conceptual data tends to favor an ultrametric space. The metric results of the color and letter datasets, which are both of perceptual nature, would be in line with this claim. Additional behavioral and neural data from other perceptual and conceptual datasets would be necessary to confirm this possibility.

A third interpretation is that the brain does contain (perceptual or conceptual) ultrametric representational spaces congruent to behavioral data, but that the behavioral-neural letter mismatch is an exceptional case. As we have suggested before (Hanson et al., 2020), this mismatch appears to be caused by a mnemonic encoding process due to the alphabet's association with the alphabet song. Alternatively, the metric scaling could be due to an artifact in the data or methodology. Next, we present multiple pieces of evidence that speak against this possibility.

Both the color and letter 2D MDS solutions produce a circular shape with linear ordering, which could be concern for a possible artifact. However, a closer examination reveals distinct geometries that seem to be characteristic of the dataset's and ROI's representations. Even though both 2D MDS solutions are circular, the MDS solutions of the color data are an open circle and accurately represent the larger gap between the wavelengths of 434nm and 674nm. Higher dimensions further reveal a figure eight shape with directional shifts of the alphabet for the letters, further strengthening the mnemonic encoding hypothesis. Meanwhile, the colors maintain their open circle geometry even in higher dimensions. Also distinct from the colors, the representations in auditory cortex resemble a helix shape. The lack of geometry in the control data of the fixation activations in auditory cortex and in the Gaussian simulations further show that the observed shapes are not a product of randomness. Furthermore, all of the geometries show consistency across subjects and within ROIs. Therefore, we believe that the behavioral-neural mismatch is caused by the mnemonic encoding of the letters in the brain. This does not exclude the possibility of ultrametric spaces in the brain, but future studies are needed to explore the possibility of a metric preference or a perceptual-conceptual dichotomy. Future research should also investigate whether mnemonically assisted learning can lead to other behavioral-neural representational discrepancies.

To our knowledge, this is the first work that extends the study of metric and ultrametric theories to the brain. Taken together, our results provide evidence that psychological spaces measured by behavior are also encoded in neural activations of the brain and that they sometimes, as in the case of colors, have equivalent representational spaces. Further investigation of the brain's representational metricity holds the exciting potential of providing new evidence to uncover the underlying cognitive processes of how object similarity is encoded.

References

- Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 Complex). *eNeuro*, 3(4), ENEURO-0039.
- Busmeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.
- Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, 21(4), 449-467.
- Ekman, G. (1954). Dimensions of color vision. *The Journal of Psychology*, 38(2), 467-474.
- Flowers, D. L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T. A., Wood, F. B., & Eden, G. F. (2004). Attention to single letters activates left extrastriate cortex. *Neuroimage*, 21(3), 829-839.
- Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive psychology*, 23(2), 222-262.
- Hahn, U., & Chater, N. (1997). Concepts and similarity. *Knowledge, concepts and categories*, 43-9
- Hanson, S. J., Caglar, L. R., & Hanson, C. (2020). Decoding Second Order Isomorphisms in the Brain: The case of colors and letters. *bioRxiv*.
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2, 4.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling* (Vol. 11). Sage.
- Kuennapas, T., & Janson, A. J. (1969). Multidimensional similarity of letters. *Perceptual and Motor Skills*, 28(1), 3-12.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological review*, 100(2), 254.
- Newton, I. (1704). *Opticks: Or a treatise of the reflexions, refractions, inflexions and colours of light*.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan*.
- Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika*, 47(1), 3-24.
- Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive psychology*, 15(3), 346-378.
- Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, 42(3), 319- 345.
- Shepard, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27(2), 125-140.
- Shepard, R. N. (1964, May). Extracting latent structure from behavioral data. In *of the 1964 symposium on digital computing*.
- Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210(4468), 390-398.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317-1323.
- Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive psychology*, 1(1), 1-17.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4), 273.
- Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4), 327.