

NEUROSCIENCE

Hyperbolic geometry of the olfactory space

Yuansheng Zhou^{1,2}, Brian H. Smith³, Tatyana O. Sharpee^{1,4*}

In the natural environment, the sense of smell, or olfaction, serves to detect toxins and judge nutritional content by taking advantage of the associations between compounds as they are created in biochemical reactions. This suggests that the nervous system can classify odors based on statistics of their co-occurrence within natural mixtures rather than from the chemical structures of the ligands themselves. We show that this statistical perspective makes it possible to map odors to points in a hyperbolic space. Hyperbolic coordinates have a long but often underappreciated history of relevance to biology. For example, these coordinates approximate the distance between species computed along dendrograms and, more generally, between points within hierarchical tree-like networks. We find that both natural odors and human perceptual descriptions of smells can be described using a three-dimensional hyperbolic space. This match in geometries can avoid distortions that would otherwise arise when mapping odors to perception.

INTRODUCTION

The reason that the sense of smell can be used to avoid poisons or estimate a food's nutrition content is because biochemical reactions create many by-products. Thus, the emission of certain sets of volatile compounds will accompany the production of a specific poison by a plant or bacteria. An animal can therefore judge the presence of poisons in the food by how the food smells. Other specific examples include the use of smell by bees when judging whether a flower has more pollen or nectar (1, 2). Fruit flies select places to lay eggs based on odors (3). These examples suggest that, from a practical perspective, it would be useful for the nervous system to classify odors based on statistics of their co-occurrence. For example, if odor components that are strongly correlated are represented nearby within the nervous system (4), then detection of one component could be quickly used as an indicator for the likely existence of another component that is strongly correlated with it. With this perspective in mind, we set out to study the structure of the olfactory space based on odor co-occurrence.

Before we describe the results, we review the reasons for why one might expect to find hyperbolic coordinates to be relevant for olfaction and biological systems in general. Biological data are often represented using dendrograms or hierarchical tree structures (Fig. 1A). These data can be equivalently represented using Venn diagrams, where larger circles correspond to broader classifications (Fig. 1B) (5). For example, before Darwin, these Venn diagrams were used to classify species based on their properties (6). Darwin used the mapping from Venn diagrams to trees (6) to infer the likely tree for speciation based on available descriptions of species properties (Venn diagrams). There is a deep mathematical reason underlying the equivalence between these two representations, and it involves hyperbolic spaces. Specifically, starting with the Venn diagram (Fig. 1B), one can assign points to a three-dimensional (3D) space whose horizontal x and y coordinates equal to center coordinates of the Venn circles, whereas the vertical coordinate equals to the circle radius (7). In this manner, larger circles get assigned to higher heights, which would then correspond to positions closer to the tip of the tree (Fig. 1C). Sometimes, the presence of partially overlapping circles leads to a structure that is not precisely a tree because it contains loop. Nevertheless, the resulting 3D space has a hyperbolic

metric (8) and can be described by the Poincare half-space model for the hyperbolic space. The fact that the metric is non-Euclidean can be observed from the fact that the shortest distance between two points goes up in the z -direction (along the tree) before descending back to the target node. In Fig. 1D, we show an example shortest path between two points in a 2D half-space model (red dashed line) and its discrete approximation (red solid line). To foreshadow the results on olfactory odor classification, we note that 3D hyperbolic space is the lowest dimensional space where the descriptor sets (Venn diagrams) are not 1D, as in Fig. 1D, but are 2D circles as in Fig. 1 (B and C). At least two axes have been described for the human odorant perception (the “pleasantness-to-unpleasantness” axis and the “chemical-to-natural” axis) (9, 10). Together, these mathematical and biological observations point to the relevance of 3D hyperbolic geometry for odor perception.

RESULTS

To analyze which space best describes the statistics of co-occurrence within natural odor mixtures, we used a recently developed a statistical

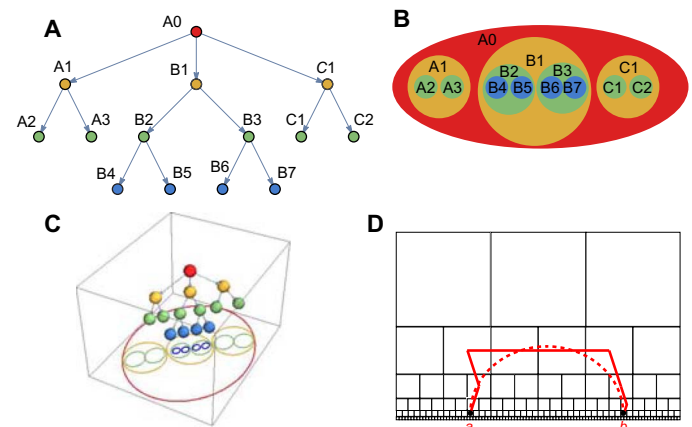


Fig. 1. Hyperbolic spaces approximate hierarchical networks. (A) Example hierarchical description of data and (B) its equivalent representation using Venn diagrams. (C) Venn diagrams can be mapped onto points in a 3D space, forming approximately a tree. The metric in the resulting 3D space is hyperbolic (7). The hyperbolic aspects of the metric are illustrated by the fact that the shortest path between nodes in the tree goes upward and then descends back to the target node. (D) Discrete approximation to a half-space model of the hyperbolic space in 2D. Red solid and dashed lines show the discrete and continuous shortest paths between point a and b within the half-space model of the hyperbolic space (8).

¹Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. ²Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA. ³School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA. ⁴Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA.

*Corresponding author. Email: sharpee@salk.edu

method (11) that can identify the presence of a geometric structure in data based on observed correlations between data components. This method is unaffected by linear or nonlinear monotonic transformations of inputs and therefore can be used to determine the overall geometry of the data without worrying at first about the precise scaling of the axes. The analysis starts by taking a set of measurements of concentrations of individual monomolecular odors, as they occur in the natural environment. Our analyses will be based on four data sets of odors measured from samples of strawberries (12), tomatoes (13), blueberries (14), and mouse urine (15). To give an overview of the data, 69 monomolecular odors were measured across 50 different mouse urine samples, 66 monomolecular odors across 79 tomato samples, 45 monomolecular odors across 101 blueberry samples, and 78 monomolecular odors across 54 samples of strawberries. The first step of the analysis is to compute correlations between the concentrations of monomolecular odor across samples (Fig. 2A). The correlation coefficient between two odors x, y was defined as

$$\text{corr}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

where x_i and y_i represent concentrations of odor x and odor y in i th sample, \bar{x} and \bar{y} represent the mean values of concentrations across all samples, and n is the number of samples. Correlation coefficients were computed separately for each data set. The absolute values of the correlation matrix from each data set are passed through a step function with different thresholds: The correlation values above the threshold are set to 1, and the rest of values are set to 0 (Fig. 2B). The transformed matrix can now be visualized as a topological graph where all unit values represent links between the corresponding odors (Fig. 2C). The graph can be characterized by the number of holes (cycles) in one, two, or higher dimensions. For high thresholds, the number of cycles will be low because most units are not connected. Similarly, at low thresholds, the number of cycles is also low because units form fully connected networks. Plotting the number of cycles as a function of density of edges, or equivalently the number of connected nodes, yields the so-called Betti curves. It turns out that the shapes of these Betti curves are quite sensitive to the statistics of correlations. This sensitivity makes it possible to infer the geometry of the space that can produce these correlations if we sample points from this space and assume that stronger correlations (before thresholding) imply closer distances (11).

Applying this statistical approach to each of the four data sets separately, we found the data in each case to be consistent with being drawn from a neighborhood of a sphere positioned within a 3D hyperbolic space together with a small amount of multiplicative noise added to the distances (Fig. 2D). The fact that hyperbolic space approximates hierarchical tree-like networks motivated this choice of the model (7), with odors reflecting leaves of the network—the neighborhood of the surface (Fig. 1). Quantitatively, one can compare Betti curves derived from a model geometrical space and from a data set by computing the integral of the curve (11), the quantity referred to as the integrated Betti value. To find the best-fitting geometry, we optimized parameters of the model such that the noise magnitude and the range of radii within the space from which the sample points were drawn provided the best match to the first integrated Betti value. Then, we examined how these optimized parameters could account for the second and third integrated Betti values. For all four data sets, we found the measurements to be consistent with sampling from a 3D

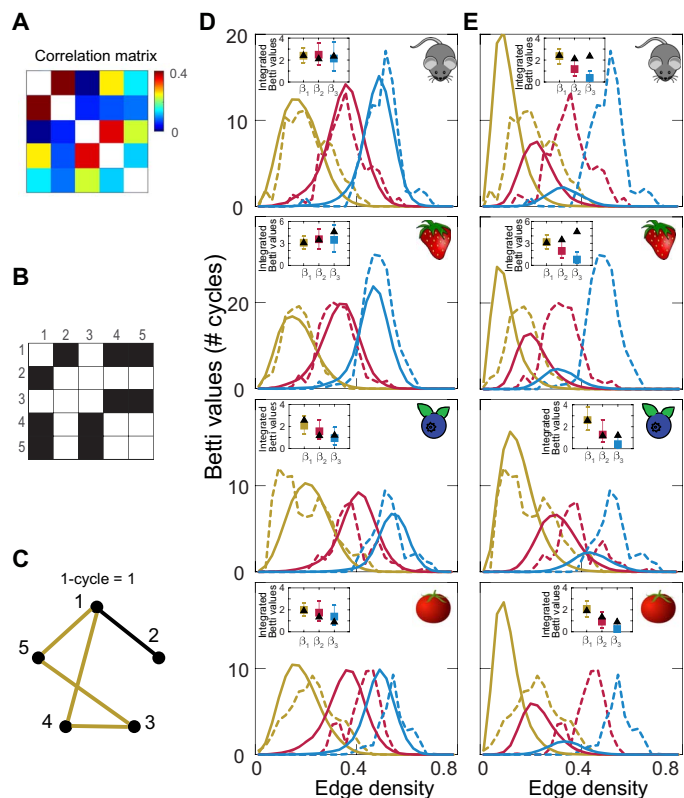


Fig. 2. Topological organization of the natural odor space. (A to C) Illustration of the topological algorithm for identifying spaces consistent with correlation statistics. (A) Example correlation matrix for five odors in strawberry data set. (B) Correlation matrix after applying a threshold of 0.25. (C) A nonzero value represents an edge connecting the two elements. The resulting complex has one 1D cycle and edge density of 0.5. (D and E) Betti curves with the number of cycles in one (yellow), two (red), and three (blue) dimensions plotted as a function of edge densities. Data from Betti curves (dashed) are compared with predictions using model geometry (solid lines) of 3D hyperbolic space in (D) or Euclidean space (E). Insets show comparisons between integrated Betti values from data (black triangles) compared with models. The error bars show 95% confidence intervals (from 2.5% to 97.5%) from 300 models with the same number of odors as the data, and the colored squares show the medium values of the models.

hyperbolic space ($P > 0.25$, $P > 0.21$, $P > 0.45$, and $P > 0.19$ for blueberry, tomato, mouse, strawberry data sets, respectively; in each case, the value stated is the minimal value across the Betti curves in 2D and 3D; see also table S1).

The first three Betti curves were also sufficient to show that Euclidean spaces could not account for the data, even when dimensionality and other parameters were optimized (Fig. 2E, $P < 0.03$ for blueberry and $P < 0.003$ for other three data sets). As a control, we verified that shuffling odor concentrations between samples, which destroys correlations between odors, produced Betti curves that can be fully explained by random matrices ($P = 0.4$, $P = 0.7$, and $P = 0.9$ for integrated Betti values one through three, respectively; cf. fig. S1). These matrices would not be consistent with the hyperbolic space plus the small amount of noise that fits the real data ($P < 0.01$). As additional controls, we verified that (i) evaluating differences between Betti curves using L1 distances instead of the integrated Betti values (tables S2 and S3) or (ii) applying logarithm to concentration values before computing their correlations led to the same conclusions (fig. S2 and

table S4). In particular, hyperbolic 3D is consistent with measurements for all three Betti curves, whereas the best-fitting Euclidean model can be ruled out according to these measures. The corresponding P values are provided in tables S1 to S4. Note that hyperbolic spaces of dimensions higher than three cannot be ruled out (fig. S3). However, the 3D hyperbolic space remains the best-fitting model across the four data sets. This is true whether one uses either the integrated Betti value or the L1 distances between model and experimental Betti curves (fig. S3).

To visualize how the points consistent with odorant correlation statistics might be distributed within the hyperbolic space, we used non-metric multidimensional scaling (MDS) (16). The nonmetric MDS algorithm embeds a set of points into the N -dimensional space while attempting to preserve the rank ordering of distances as best as possible (16). Traditionally, MDS is applied to the Euclidean space, but we modified it (see Materials and Methods) to work with hyperbolic distances (7). After testing the algorithm on synthetic data (fig. S4), we applied the modified algorithm to the four data sets. In Fig. 3, we show results for the four data sets. Because the points are located near a surface of a sphere (the range of radii $R_{\min} = 0.9 R_{\max}$), we present the points on a sphere using the two angles of latitude and longitude. The results show approximately uniform sampling in all four data sets. Notably, the points do not cluster based on functional chemical properties of the individual components (fig. S5). One can understand the absence of clustering from the fact that monomolecular odors with different functional properties are produced together in biochemical pathways.

What are the axes of this olfactory space? Or, in other words, do odors associated with certain parts of this space have different perceptual or physicochemical properties? Previous studies found axes that correlated with perceptual odor pleasantness (17) and physicochemical properties such as molecular boiling point and acidity (9, 17, 18). We checked for

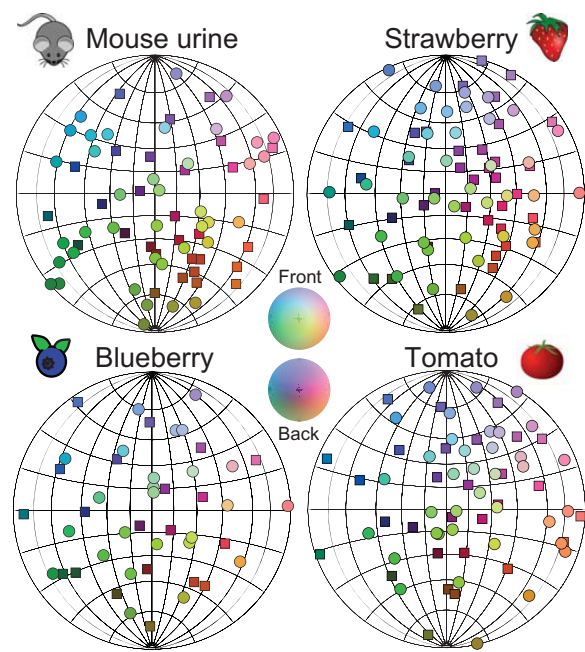


Fig. 3. Visualization of the natural olfactory space using nonmetric MDS. Because the variation in radius is small, data points are shown on the surface of a sphere with circles/rectangles for points falling on the near/far side of the sphere. The RGB color scales were proportional to the XYZ coordinates of points.

associations with all of these properties, and supporting previous findings (9, 10, 17, 19) found that points corresponding to pleasant and unpleasant odors occupied different parts of the space. We note that this analysis of pleasantness rankings was based only on odor components from tomato and strawberry data sets, for which these rankings were available. Thus, the pleasant-unpleasant odor axis can be identified even solely using fruit odor components. The direction most associated with a change in pleasantness value is marked by the red line in Fig. 4 (A and C). For odor mixtures produced by individual fruit samples, we use the “overall liking” rating assigned by humans to fruit samples as a measure of pleasantness (Fig. 4A). To test how well the identified pleasantness axes can predict measured pleasantness rankings for novel samples, we regenerated this axis using only strawberry samples and use it to predict pleasantness ranking for tomato samples. The correlation was significant, with correlation coefficient $R = 0.34$ and $P = 0.01$ (Fig. 4B). The pleasantness values could also be assigned to individual odor components based on the correlation between the odor concentration in a mixture and mixture pleasantness computed over samples (Fig. 4C). This measure of pleasantness produced an even stronger correlation between pleasantness and odor coordinates within the space, $R = 0.66$ and $P = 3 \times 10^{-7}$ (Fig. 4D). We also computed these correlation values using different odor components from those used to generate the pleasant-unpleasant axis for odor components in Fig. 4C. Specifically, we used two-thirds of randomly selected odor components to generate this axis; we computed the correlation value using the remaining components. The pleasantness axis for odor components had a similar orientation to the one for mixtures (that is, individual fruit samples).

In addition to the pleasantness axis, we could also find axes that were strongly associated with two other properties: molecular boiling point—which is probably a reflection of volatility—and acidity, both of which showed significant correlations ($P < 0.04$; Fig. 4, E and F). We assigned acidity for individual odors as the correlation coefficient between its concentration and fruit sample acidity measurement. (We computed all of these correlation coefficients using a different subset of odors from the one used to estimate the corresponding axes.) Because the space is essentially 2D, the three axes of odor pleasantness, acidity, and molecular boiling point are not independent. In other words, knowing the coordinates along the molecular boiling point and acidity axes, one can predict the position along the pleasantness axis. That is, the identified mapping to a sphere in a hyperbolic space makes it possible to predict, with correlation $R = 0.34$ (Fig. 4B) for natural mixtures and with $R = 0.66$ (Fig. 4D) for monomolecular odors, how perceptually pleasant these odors are based on their projections on the acidity and volatility axes.

The observation that the odor mixtures can be mapped onto a continuous metric space is consistent with a previous vector-based model of human olfactory perception (19). This model posits that the perception of odor mixtures is based on a combination of the mixture components or, in other words, that there is an underlying set of coordinates that can represent olfactory mixtures. Previous analysis (9) of the Dravnieks database (20) containing human perceptual descriptions of >120 monomolecular odors showed that the perceptual space is likely to be curved. Qualitatively, the points were found to form a “potato-chip” surface (9). This can be a signature of the hyperbolic space; potato-chip or saddle-like surface have a negative curvature and serve as an everyday example of hyperboloid surfaces. To quantitatively test whether the perceptual space is described by hyperbolic geometry, we applied the Betti curve method to the Dravnieks database (20). First, we found that Euclidean spaces were not consistent with measured Betti curves ($P < 0.003$;

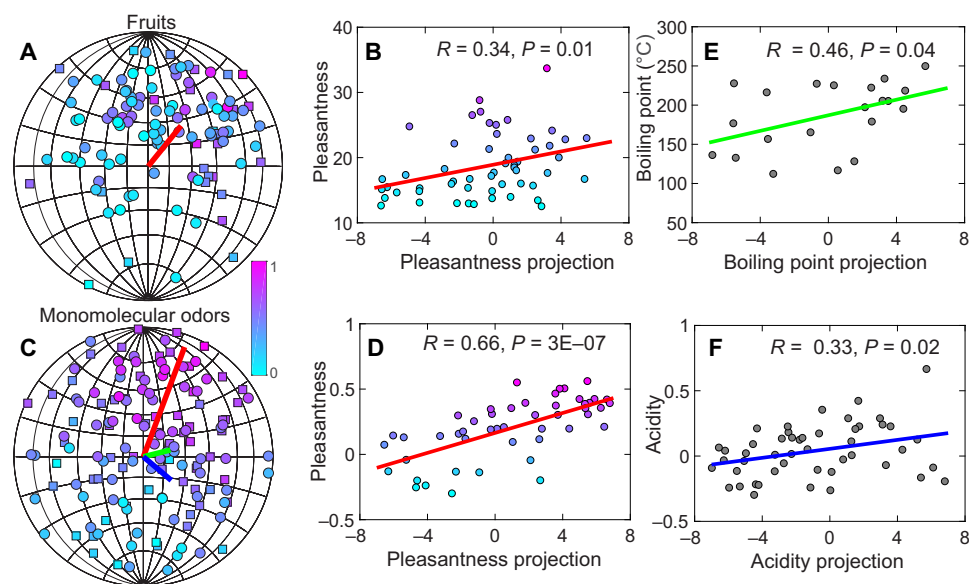


Fig. 4. Axes associated with pleasantness and odor physicochemical properties. (A) We represented 108 fruit samples (54 tomatoes and 54 strawberries) using normalized linear combinations of odor coordinates in the space, with the weights proportional to odor concentrations in the samples. The color indicates human rating for the overall liking; circles/squares represent points from the front/back sides of the sphere. The red line shows the direction most associated with the pleasantness ratings. (B) The correlation is significant, with correlation value $R = 0.34$ and $P = 0.01$. (C) Visualization of 144 individual monomolecular odors. The red, green, and blue lines showed the directions of pleasantness, boiling point, and acidity, respectively. Two-thirds of monomolecular odors were used to determine the directions associated with perceptual or physicochemical properties of odors, and the rest one-third were projected onto the directions as validation sets to evaluate the correlations. In the 144 monomolecular odors, only 62 of them have available boiling points and were used to find the boiling points direction. (D) Correlation between odor pleasantness with projection onto the pleasantness axes. (E) Correlation between molecular boiling point, a measure of odor volatility, with projection on the associated axes. (F) Correlation between acidity value and the associated axes.

Fig. 5A and table S5). The first Betti curve could not be matched to the data in terms of its area for any dimensionality of the Euclidean space (Fig. 5A, inset). Second, we found that the full hyperbolic space of varying dimensions could match the area of the first Betti curve. However, only hyperbolic spaces with small number of dimensions could also simultaneously match the area of the second Betti curve (Fig. 5B, inset). The 3D hyperbolic space produced the best fit, with larger dimensions yielding increasing deviations. Hyperbolic spaces with dimensions nine and above could be excluded with $P < 0.034$. The third Betti curve was essentially zero and is not shown here. One may notice that the first and second Betti curves were not as regular as in the case of odorants and contained multiple peaks. It turns out that the biphasic nature of the Betti curve could be explained by the nonuniform distribution of points across the two angles (Fig. 5C). Unlike in the case of olfactory stimulus spaces that are sampled approximately uniformly, here, the distribution of points obtained using MDS is not uniform and clusters in one-half of the space. Sampling points from this embedding yields biphasic Betti curves that match those derived from perceptual data (Fig. 5D). Specifically, P values for L1 differences between Betti curves derived from data and MDS fits were $P = 0.32$ (hyperbolic, Betti 1), $P = 0.20$ (hyperbolic, Betti 2), $P = 0$ (Euclidean, Betti 1), and $P = 0.06$ (Euclidean, Betti 2) (cf. inset in Fig. 5D). The MDS distances also better correlated with perceptual distances when we carried out MDS in the hyperbolic space compared to Euclidean space (fig. S6).

DISCUSSION

Our results highlight the importance of hyperbolic curved geometry for understanding how natural odors are represented in the nervous system. Overall, we find that both the statistics of natural odor mix-

tures and human odor perception can be mapped onto hyperbolic spaces. In the natural environment, hierarchical biochemical networks produce odor components. Hierarchical networks can often be approximated by trees and, therefore, by hyperbolic spaces (7, 21). We find that most natural odor components fall near the boundary of the observed hyperbolic space, corresponding to leaves of the trees (Fig. 1). At the perceptual level, we also found hyperbolic organization. However, in this case, the odors selected for the Dravnieks database did not sample the human perceptual space uniformly (Fig. 5).

Hyperbolic perceptual organization is likely to be general across different sensory modalities. There are two reasons for this. First, neural networks that give rise to perception are hierarchically organized, and as we have seen in Fig. 1, this can lead to hyperbolic geometry. Second, individual neurons have limited response ranges. Because of response saturation, small changes in neural responses near their limit correspond to exponentially large changes in the input values. This compressive mapping (22) is similar to the Poincaré disk representation of the hyperbolic space (7). There is evidence that visual, haptic, and auditory perceptual spaces are all hyperbolic (23–26). Adding olfactory perception to this list could help explain why humans can map odors to auditory pitch (27, 28) and to colors (29).

Noteworthy is the low dimensionality of both the physical odor space and perceptual odor spaces. In both cases, the curved space contains approximately three dimensions despite the fact that the data vary in >50 dimensions associated with different samples of natural odor mixtures and according >100 perceptual descriptors. The low dimensionality of the environmental odor space could be a general property of natural odors because it occurred for odors as diverse as fruit and mammalian urine odors. Note that all four natural odor data sets were described by the same 3D hyperbolic space with exactly the

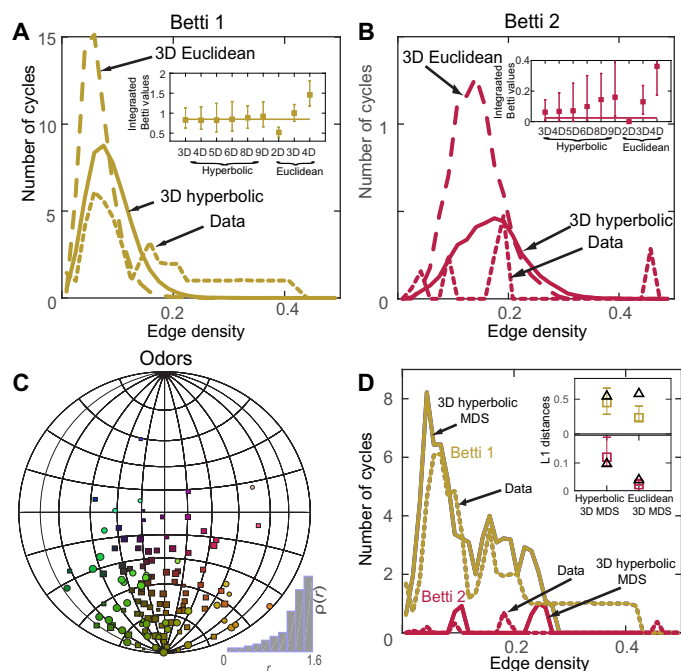


Fig. 5. Hyperbolic organization of the human olfactory perception. The (A) first and (B) second Betti curves of the perceptual data set (dotted line) (20) compared to Betti curves of the 3D hyperbolic space (solid line) and 3D Euclidean space (dashed line). Euclidean and hyperbolic spaces of other dimensions provided a worse fit. Insets compare integrated Betti values from data (horizontal lines) and 300 repeated models in different dimensions with Euclidean or hyperbolic metrics. The error bars show 95% confidence intervals; the number of repeated computations of model curves was 300. (C) Visualization of odors in human olfactory perception space using nonmetric MDS in a 3D hyperbolic space. The sizes of points are proportional to their radii. The radius distribution is shown in bottom right inset. (D) The multimodal aspects of Betti curves derived from data (dotted lines) can be accounted for by the nonuniform distribution of points within the 3D hyperbolic space. Sampling points from (C) produces multimodal first (yellow) and second (red) Betti curves (solid lines). Inset shows comparison of L1 distances between Betti curves derived from data and those derived from 100 different MDS fits. Black open triangles represent the distance between data and model mean, and colored bar plots show the range of values, where data curves are substituted by different MDS fits.

same radius (equivalent to curvature of the space). This property could make it easier to represent data from different data sets within the same space. For example, odors from strawberry and tomato could be represented jointly within a single 3D space (Fig. 4). We could not combine data from other data sets because, for example, there were no overlapping components between fruit odors and mouse urine data sets. It is possible that representing all possible natural odors will increase the dimensionality of the overall space. Another possibility is that introducing odors from different sources will “fill in” the inner part of the hyperbolic space. The natural odors considered here mapped onto a surface of a hyperbolic space. Odors produced by biochemical pathways of different complexity are likely to map to surfaces with a different radius, filling in the space. This possibility is especially interesting because it would provide a link to the filled 3D hyperbolic space that we find for perceptual data, which was obtained using diverse classes of odors. At the same time, the perceptual odor mapping reveals that odors tested so far concentrate on one side of the space (Fig. 5) (9, 19), whereas natural odor components cover their respective space rather uniformly (Fig. 3). These analyses thus suggest perceptual coordinates that are yet to be explored.

The match in dimensionality between the environmental and perceptual spaces would not have been expected a priori. The matching dimensionality between the input and perceptual spaces can help avoid nonlinear distortions that would necessarily arise when mapping two nonlinear spaces of different dimensionality. These distortions are known to exist in vision where we perceive distances in a compressed way: The moon appears disproportionately closer to us than would be based on the actual Euclidean distance (23, 24). We also plot equidistant and parallel lines differently, which is one of the key signatures of the hyperbolic space. Similar distortions arise in the haptic space (25). The matching geometry between the input and perceptual spaces in olfaction may therefore serve to minimize these distortions in odor perception. Overall, the ability of the perceptual system to resolve points in the low-dimensional odor space would depend on the number and tuning properties of sensory receptors (30–34).

MATERIALS AND METHODS
A clique topology method for finding geometric spaces consistent with correlations in the data

We followed procedures from (11) to generate Betti curves for samples taken from spaces with different geometries. The method effectively converts the correlation matrix to its rank-ordered version. This renders the algorithm’s results invariant under monotonic transformation of values, for example, due to nonlinearities introduced at the measurement stage. However, this property can also be used to assign a distance between points based on the correlation in the activity of two units in a network (11) or, as in our case, between two odors across different samples. All monotonic functions will yield the same result. We chose $D_{ij} = -|C_{ij}|$, where D_{ij} is the assigned distance between odors, and $|C_{ij}|$ is the absolute value of the correlation coefficient of odor concentrations among a set of points. This definition ensures that stronger correlations (in absolute value) corresponded to tighter connections and smaller geometric distances, as in (35).

The first three Betti curves turn out to be quite sensitive measures of the distance matrices and can be used to find underlying geometries consistent with the data (11). In addition to random spaces, we screened two kinds of geometric structures: Euclidean spaces of different dimensionality and hyperbolic space [we used the hyperbolic ball model (7) with curvature $\zeta = 1$] with different parameters. In each space, we uniformly sampled points (the same number as the number of odors in each of the data sets) based on the metric of the space. In a d -dimensional Euclidean space, the points were uniformly distributed in a d -cube with Euclidean distance. For a d -dimensional hyperbolic ball model, we used partial space by setting the minimal radius R_{min} and maximal radius R_{max} for the ball. This choice of the model was motivated by the fact that hyperbolic space approximates hierarchical tree-like networks (7), with odors reflecting leaves—the neighborhood of the surface. We took the angles of points randomly and selected radii r within $[R_{min}, R_{max}]$ following the distribution

$$\rho(r) \sim \sinh((d - 1)r) \tag{1}$$

The distance between two points was derived from hyperbolic law of cosines

$$\cosh \zeta x = \cosh(\zeta r) \cosh(\zeta r') - \sinh(\zeta r) \sinh(\zeta r') \cos \Delta \theta \tag{2}$$

where ζ is the curvature set as 1 in our model, r and r' are the radial distances of the two points, and $\Delta\theta$ is the angle between them. Considering that noise may exist in the monotonous correspondence between the underlying topological distance and correlation strength of odors and that the amount of noise may differ between data sets due to differences in sample collection procedures, we added multiplicative Gaussian noise to the distance matrices for both Euclidean model and hyperbolic model before plotting the Betti curves. Together, we have the topological distance matrices of the sampled points in geometric spaces

$$D = D_{\text{geo}} \cdot (1 + \varepsilon \cdot N(0, 1)) \quad (3)$$

where D_{geo} is the geometric distances, and ε is the noise level.

In summary, the space geometry affects Betti curves through the distribution of sampled point density (Eq. 1) and distance measures (Eq. 2). Multiplicative noise (Eq. 3) also affects Betti curves. The optimal parameter values for the hyperbolic model were $R_{\text{max}} = 7$ and $R_{\text{min}} = 0.9 R_{\text{max}}$ for all four odor data sets, while optimal noise values were $\varepsilon = 0.045$ (mouse), $\varepsilon = 0.050$ (strawberry), $\varepsilon = 0.050$ (blueberry), and $\varepsilon = 0.040$ (tomato). The optimized parameters for the Euclidean space were as follows: mouse data set, dimension $d = 8$; noise, 0.05; strawberry data set, dimension $d = 10$; noise, 0.05; blueberry data set, dimension $d = 10$; noise, 0.09; tomato data set, dimension $d = 8$; noise, 0.09.

For the perceptual data set (127 monomolecular odors, 146 descriptions for each) (20), the topological pairwise distances of odors k and n were defined as $\sum_{i=1}^{146} (v_i^k - v_i^n)^2$, where v_i^k denotes human descriptions for the k th odor. We use the differences between descriptions across odors because, in this case, the absolute value of the descriptor matters, unlike in the case of odors where correlations were a more appropriate measure. When fitting the data using geometric models, no noise was added to distances in models. We also tested the sensitivity of the Betti curves to noise in pairwise perceptual distances between odors. This was carried out by computing perceptual distances based on randomly selected subset of 120 of the total 146 descriptors. The variability in the resultant distance values was proportional to the mean distance (fig. S7). The relative error in the integrated Betti values across these samples was the same as the relative error of the distances themselves (fig. S7, inset on the right). Thus, although the Betti curve construction evaluates data structure globally, it is not driven by variability in larger distances. In the case of the perceptual data set, we found that the full hyperbolic space better described the data rather than a shell, and therefore, the minimal radius was set to zero. We optimized maximal radius of the hyperbolic model, which is a measure of its curvature, to fit the integrated Betti value of the first Betti curve. The optimal R_{max} were as follows: 1.6 (3D), 1.9 (4D), 1.8 (5D), 1.7 (6D), 1.9 (8D), and 3.0 (9D). We used these values to compute the second Betti curve and determine how well it could account for the second integrated Betti value.

All reported P values for comparison with experimentally generated Betti curves were obtained by creating, for each candidate geometry, 300 statistically equivalent models. Points for each model were selected randomly according to the density specific to that geometry (uniformly within the unit cube for Euclidean spaces and according to Eq. 1 for hyperbolic spaces). The number of points was matched to the number of points in the corresponding experimental data set. On the basis of this simulated point distribution, we computed 300 different

Betti curves. These curves were then used to generate a distribution of integrated Betti values or compute the L1 distance of these curves from the mean Betti curve of this model. The reported P values reflect two-tailed percentiles for where experimental Betti curves fall within the model-generated distributions. We report P values as <0.003 when none of the samples generated values further from the mean than the observed data point.

Nonmetric MDS embedding of odors on the surface of a 3D hyperbolic sphere

The nonmetric MDS algorithm embeds a set of points within a prespecified space while attempting to preserve rank-ordered distances between points. We modified the Euclidean-based, nonmetric MDS algorithm in MATLAB version 2017a by replacing the Euclidean distance with hyperbolic distance in Eq. 2. The initial positions of points were uniformly sampled in the optimal 3D hyperbolic space determined in Fig. 1. The radial coordinates were fixed because their range was small and points were approximately positioned on the surface of a sphere. The algorithm updated the angular coordinates to minimize the mismatch in the rank order of distances. The iterations ended when this error fell below a threshold of 0.001. Because the MDS algorithm can return the arbitrary rotation of the space, in Fig. 3, we used the Procrustes algorithm to align the positions of odors between the strawberry and tomato data sets, using the strawberry data set that had the most odors as an anchor. The Procrustes process was carried out through the Procrustes function in MATLAB, and the scale component and translation component were set to 1 and 0, respectively.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/8/eaag1458/DC1>

Fig. S1. No indications of hyperbolic geometry in shuffled odor data sets.

Fig. S2. Alternative ways of evaluating differences between Betti curves also support hyperbolic geometry of natural odor spaces.

Fig. S3. Error bar plots of Betti curves statistics for the hyperbolic model of different dimensions.

Fig. S4. Test of the nonmetric multidimensional scaling algorithm in the hyperbolic space on synthetic data.

Fig. S5. Odors within the identified space do not cluster by functional group.

Fig. S6. Comparison between embedded geometric distances and reported perceptual distances.

Fig. S7. Analysis of sensitivity of integrated Betti value to noise in the input distances.

Table S1. Statistical tests (P values) for consistency with hyperbolic models based on integrated Betti values.

Table S2. Statistical tests (P values) for consistency with hyperbolic models based on L1 distances between Betti curves.

Table S3. Statistical tests (P values) for evaluating consistency of experimental Betti curves with respect to 3D hyperbolic model or optimal optimal Euclidean model.

Table S4. Statistical tests (P values) for evaluating consistency of Betti curves computed based on logarithm of odor concentrations with respect to hyperbolic model.

Table S5. P values of hyperbolic and Euclidean model using integrated Betti values for perceptual data set.

REFERENCES AND NOTES

1. H. E. M. Dobson, E. M. Danielson, I. D. Van Wesep, Pollen odor chemicals as modulators of bumble bee foraging on *Rosa rugosa* Thunb. (Rosaceae). *Plant Species Biol.* **14**, 153–166 (1999).
2. A. C. Knauer, F. P. Schiestl, Bees use honest floral signals as indicators of reward when visiting flowers. *Ecol. Lett.* **18**, 135–143 (2014).
3. R. Azanchi, K. R. Kaun, U. Heberlein, Competing dopamine neurons drive oviposition choice for ethanol in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 21153–21158 (2013).
4. H. B. Barlow, Why have multiple cortical areas? *Vision Res.* **26**, 81–90 (1986).

5. B. Johnson, B. Shneiderman, Tree-maps: A space-filling approach to the visualization of hierarchical information structures, in *Proceedings of the 2nd International IEEE Visualization Conference* (IEEE, 1991), pp. 284–291.
6. M. A. Ragan, Trees and networks before and after Darwin. *Biol. Direct* **4**, 43 (2009).
7. D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, M. Boguñá, Hyperbolic geometry of complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **82**, 036106 (2010).
8. J. W. Cannon, W. J. Floyd, J. Kenyon, W. Parry, Hyperbolic geometry, in *Flavors of Geometry* (MSRI Publications, 1997), vol. 31, pp. 59–115.
9. A. A. Koulakov, B. E. Kolterman, A. G. Enikolopov, D. Rinberg, In search of the structure of human olfactory space. *Front. Syst. Neurosci.* **5**, 65 (2011).
10. E. C. Crocker, L. F. Herderson, Analysis and classification of odors. An effort to develop a workable method. *Amer. Perfum.* **22**, 325–356 (1927).
11. C. Giusti, E. Pastalkova, C. Curto, V. Itskov, Clique topology reveals intrinsic geometric structure in neural correlations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13455–13460 (2015).
12. M. L. Schwieterman, T. A. Colquhoun, E. A. Jaworski, L. M. Bartoshuk, J. L. Gilbert, D. M. Tieman, A. Z. Odabasi, H. R. Moskowitz, K. M. Folta, H. J. Klee, C. A. Sims, V. M. Whitaker, D. G. Clark, Strawberry flavor: Diverse chemical compositions, a seasonal influence, and effects on sensory perception. *PLOS ONE* **9**, e88446 (2014).
13. D. Tieman, P. Bliss, L. M. McIntyre, A. Blandon-Ubeda, D. Bies, A. Z. Odabasi, G. R. Rodriguez, E. van der Knaap, M. G. Taylor, C. Goulet, M. H. Mageroy, D. J. Snyder, T. Colquhoun, H. Moskowitz, D. G. Clark, C. Sims, L. Bartoshuk, H. J. Klee, The chemical interactions underlying tomato flavor preferences. *Curr. Biol.* **22**, 1035–1039 (2012).
14. J. L. Gilbert, M. J. Guthart, S. A. Gezan, M. Pisaroglo de Carvalho, M. L. Schwieterman, T. A. Colquhoun, L. M. Bartoshuk, C. A. Sims, D. G. Clark, J. W. Olmstead, Identifying breeding priorities for blueberry flavor using biochemical, sensory, and genotype by environment analyses. *PLOS ONE* **10**, e0138494 (2015).
15. F. Röck, K. P. Haderl, H. G. Rammensee, P. Overath, Quantitative analysis of mouse urine volatiles: In search of MHC-dependent differences. *PLOS ONE* **2**, e429 (2007).
16. I. Borg, P. J. F. Groenen, *Modern Multidimensional Scaling Theory and Applications* (Springer, ed. 2, 2005).
17. R. M. Khan, C. H. Luk, A. Flinker, A. Aggarwal, H. Lapid, R. Haddad, N. Sobel, Predicting odor pleasantness from odorant structure: Pleasantness as a reflection of the physical world. *J. Neurosci.* **27**, 10015–10023 (2007).
18. M. Zarzo, Underlying dimensions in the descriptive space of perfumery odors: Part II. *Food Qual. Prefer.* **43**, 79–87 (2015).
19. K. Snitz, A. Yablonka, T. Weiss, I. Frumin, R. M. Khan, N. Sobel, Predicting odor perceptual similarity from odor structure. *PLOS Comput. Biol.* **9**, e1003184 (2013).
20. A. Dravnieks, ASTM Committee E-18 on Sensory Evaluation of Materials and Products, Section E-18.04.12 on Odor Profiling, *Atlas of Odor Character Profiles* (American Society for Testing and Materials, 1985).
21. M. Boguñá, F. Papadopoulos, D. Krioukov, Sustaining the Internet with hyperbolic mapping. *Nat. Commun.* **1**, 62 (2010).
22. J. Berkowitz, T. Sharpee, Decoding neural responses with minimal information loss. *bioRxiv* 10.1101/273854 (2018).
23. T. Indow, *The Global Structure of Visual Space* (World Scientific, 2004), 216 pp.
24. R. K. Luneburg, *Mathematical Analysis of Binocular Vision* (The Dartmouth Eye Institute, 1947), 104 pp.
25. W. Blumenfeld, The relationship between the optical and haptic construction of space. *Acta Psychol.* **2**, 125–174 (1937).
26. D. Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice* (Oxford Univ. Press, 2011).
27. K. Belkin, R. Martin, S. E. Kemp, A. N. Gilbert, Auditory pitch as a perceptual analogue to odor quality. *Psychol. Sci.* **8**, 340–342 (1997).
28. O. Deroy, A. S. Crisinel, C. Spence, Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychon. Bull. Rev.* **20**, 878–896 (2013).
29. A. N. Gilbert, R. Martin, S. E. Kemp, Cross-modal correspondence between vision and olfaction: The color of smells. *Am. J. Psychol.* **109**, 335–351 (1996).
30. M. Meister, On the dimensionality of odor space. *eLife* **4**, e07865 (2015).
31. C. Bushdid, M. O. Magnasco, L. B. Vosshall, A. Keller, Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).
32. Y. Zhang, T. O. Sharpee, A robust feedforward model of the olfactory system. *PLOS Comput. Biol.* **12**, e1004850 (2016).
33. D. Zwicker, A. Murugan, M. P. Brenner, Receptor arrays optimized for natural odor statistics. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5570–5575 (2016).
34. A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, L. F. Abbott, Optimal degrees of synaptic connectivity. *Neuron* **93**, 1153–1164 (2017).
35. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article 17 (2005).

Acknowledgments: We thank V. Itskov, A. Samuel, and E. Hong for the helpful discussions and D. Kastner, V. Lundblad, and C. G. Galizia for the comments on the manuscript. **Funding:** This research was supported by the Aileen Andrew Foundation and NSF award numbers IIS-1254123, IOS-1724421, and IOS-1556388 to T.O.S. and 1556337 to B.H.S. The last two awards resulted from an NSF Ideas Lab “Cracking the Olfactory Code.” **Author contributions:** All authors participated in the design of this study and writing of the manuscript. Y.Z. analyzed the data. T.O.S. motivated the relevance of hyperbolic spaces for this project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data and codes related to this paper may be requested from the authors. Original data sets can be found at the following addresses: <https://doi.org/10.1371/journal.pone.0088446.s005> (strawberry); <https://ars.els-cdn.com/content/image/1-s2.0-S0960982212004083-mm2.zip> (tomato); <https://doi.org/10.1371/journal.pone.0138494.s002> (blueberry); <http://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0000429.s005> (mouse urine). Odors for which most of measured concentration values were zero were removed before the analysis.

Submitted 5 October 2017

Accepted 19 July 2018

Published 29 August 2018

10.1126/sciadv.aaq1458

Citation: Y. Zhou, B. H. Smith, T. O. Sharpee, Hyperbolic geometry of the olfactory space. *Sci. Adv.* **4**, eaaq1458 (2018).