

1 **Pairwise Relative Distance (PRED) is an intuitive and robust**  
2 **metric for assessing vector similarity and class separability**

3

Aarush Mohit Mittal<sup>1</sup>, Andrew C. Lin<sup>2</sup>, Nitin Gupta<sup>1,3,\*</sup>

<sup>1</sup> Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

<sup>2</sup> Department of Biomedical Science, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK

<sup>3</sup> Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

\* Correspondence: [guptan@iitk.ac.in](mailto:guptan@iitk.ac.in)

4

5 ORCID IDs:

6

7 A.M.M. 0000-0002-8633-3126

8

9 A.C.L. 0000-0001-6310-9765

10

11 N.G. 0000-0002-8408-3848

12

## 13 **Abstract**

14 Scientific studies often require assessment of similarity between ordered sets of values. Each  
15 set, containing one value for every dimension or class of data, can be conveniently  
16 represented as a vector. The commonly used metrics for vector similarity include angle-based  
17 metrics, such as cosine similarity or Pearson correlation, which compare the relative patterns  
18 of values, and distance-based metrics, such as the Euclidean distance, which compare the  
19 magnitudes of values. Here we evaluate a newly proposed metric, pairwise relative distance  
20 (PRED), which considers both relative patterns and magnitudes to provide a single measure  
21 of vector similarity. PRED essentially reveals whether the vectors are so similar that their  
22 values across the classes are separable. By comparing PRED to other common metrics in a  
23 variety of applications, we show that PRED provides a stable chance level irrespective of the  
24 number of classes, is invariant to global translation and scaling operations on data, has high  
25 dynamic range and low variability in handling noisy data, and can handle multi-dimensional  
26 data, as in the case of vectors containing temporal or population responses for each class. We  
27 also found that PRED can be adapted to function as a reliable metric of class separability  
28 even for datasets that lack the vector structure and simply contain multiple values for each  
29 class.

30

## 31 **Introduction**

32 Vectors are ubiquitous data structures. As a result, the assessment of vector similarity is one  
33 of the most frequently performed data operations in diverse areas of science and engineering.  
34 To list examples within only biology, vector similarity has been used to show that reef fish  
35 species in different ecoregions resemble each other in traits, not taxonomy or phylogeny  
36 (McLean et al., 2021); that cancerous cell lines' gene expression patterns cluster according to  
37 their tissue of origin and cancer stage (Ross et al., 2000); and that certain brain regions have  
38 similar fMRI brain activation patterns over time, suggesting they are functionally connected  
39 (Sasai et al., 2021). In these examples, the vectors represented the trait, taxonomical or  
40 phylogenetic properties of each ecoregion; the gene expression profile of each cell line; and  
41 the temporal activation pattern of each brain region, respectively. Similarly, other examples  
42 of scientific data that can be represented as vectors include the firing rates of a cortical  
43 neuron to different visual stimuli (Hubel and Wiesel, 1962; Stringer et al., 2019; Victor and  
44 Purpura, 1996), the eye blinking rates of a human under different airflow conditions  
45 (VanderWerf et al., 2003), and the sensory preferences of an animal to a given stimulus at  
46 different time points (Buchanan et al., 2015; Honegger et al., 2020; Kain et al., 2015;  
47 Linneweber et al., 2020)). Any scientific question involving the comparison of such vectors  
48 requires metrics that can determine the level of similarity between vectors.

49 Common metrics for vector similarity include Pearson's correlation, cosine similarity, and  
50 Euclidean distance. Distance-based metrics, like Euclidean distance or Manhattan distance,  
51 compare the magnitude of difference between the values in the two vectors. On the other

52 hand, angle-based metrics, like the cosine similarity or the Pearson's correlation, compare the  
53 relative pattern of values within a vector with that in another vector. To take a straightforward  
54 example, consider the vectors [1 2 3] and [10 20 30]. A distance-based metric would call  
55 them different, while an angle-based metric would call them very similar. On the other hand,  
56 the vectors [1 2 3] and [3 2 1] would be described as relatively similar by the distance-based  
57 metrics and dissimilar by the angle-based metrics. Both types of metrics provide useful and  
58 complementary information; however, in practice, multiple metrics are rarely used together.  
59 In many applications, instead of choosing between one of the two types of metrics, it would  
60 be desirable to combine the similarity in the magnitudes and the similarity in the relative  
61 patterns into a single, reliable indicator of vector similarity.

62 We recently devised a metric, called Pairwise Relative Distance (PRED), to quantify the level  
63 of similarity in different individuals' neuronal responses to the same set of odors (Mittal et  
64 al., 2020). PRED captured the similarities both in the absolute values and the across-odor  
65 patterns of the responses and provided more intuitive values of similarity than correlation in  
66 quantifying stereotypy in sensory responses (Mittal et al., 2020). These initial results led us to  
67 ask whether PRED could serve as a general-purpose metric for analyzing vector similarity in  
68 different types of datasets.

69 Here, we generalize PRED as a robust metric for assessing vector similarity and class  
70 separability. Using simulations and experimental data, we show the advantages of PRED over  
71 the commonly used metrics and demonstrate its reliability in analyzing noisy or incomplete  
72 data. We illustrate PRED's ability to capture the similarity in temporal or population-level  
73 data while preserving the dataset's structure. Although we illustrate the usefulness of PRED  
74 using examples from the olfactory system, one can use PRED equally well in other sensory  
75 modalities in neuroscience, non-neuroscience biological fields like the examples described  
76 above, and non-biological fields like machine learning. Overall, our results present Pairwise  
77 Relative Distance as a reliable metric of similarity or separability in neuroscience and  
78 beyond.

79

## 80 **Results**

### 81 **PRED as a general metric for vector similarity**

82 In this work, we generalize PRED to all datasets that can be expressed as a matrix, whose  
83 columns are specific classes (dimensions) and rows are the vectors being compared; we will  
84 refer to this organization as class-vector structure (**Figure 1a**). For example, consider the  
85 responses of different retinal neurons to the same set of visual stimuli. In this case, each  
86 visual stimulus can be considered a class (column) and each neuron (row) a vector of  
87 responses to the different classes (i.e., the set of stimuli). For any such dataset, PRED  
88 provides a unified measure of the similarity between the vectors and the separability of the  
89 classes. Put simply, class-vector PRED measures whether vector A's value in a class is more  
90 similar to vector B's value in the same class than to B's value in another class. PRED is high

91 when the distances are larger between values belonging to different vectors and different  
92 classes than between values belonging to different vectors but the same class (**Figure 1a**). In  
93 other words, a high value of PRED means that the two vectors have values not only with  
94 similar magnitudes but also with similar patterns across the classes. A zero value of PRED  
95 indicates that the two vectors have unrelated patterns across the classes. A negative value of  
96 PRED indicates that the two vectors have opposite patterns across the classes. Unlike  
97 correlation, PRED also accounts for the absolute differences between the values in the given  
98 vectors.

99 We compared PRED and five other metrics on their ability to report the similarity across  
100 vectors within a class-vector dataset. These five metrics included Pearson's correlation (PC),  
101 Cosine similarity (COS), Manhattan distance (MAN), Euclidean distance (EUC), and  
102 Chebyshev's distance (CHEB). PRED, PC, and COS values range between -1 and 1, where 1  
103 denotes high similarity; MAN, EUC, and CHEB range from 0 to  $\infty$ , where 0 denotes high  
104 similarity. To enable a direct comparison of the values of all these metrics, we transformed  
105 the distance-based metrics (MAN, EUC, and CHEB) to a range between 0 and 1 using a  
106 negative exponential (see **Materials and Methods**), such that 1 denotes high similarity for all  
107 the metrics (**Supplementary Figure 1a (i)**). We use the transformed distance-based metrics  
108 in all subsequent analyses unless otherwise stated.

109 For interpreting the values of a metric, it is helpful to know its chance level, i.e., the metric's  
110 expected value for random data. For example, suppose a metric's observed value for a given  
111 dataset is high relative to its chance level. In that case, one can reasonably infer that the  
112 vectors in the dataset have a high similarity: the more the difference, the higher the similarity.  
113 It is further desirable that the chance level remains unchanged with the size of the dataset (the  
114 number of classes in the dataset) so that values obtained from different datasets, regardless of  
115 their size, can be directly compared. To test each metric's chance level, we simulated two  
116 different random datasets, one with 2 and the other with 5 classes. Each dataset included 10  
117 vectors (with length equal to the number of classes) sampled from a uniform distribution  
118 between 0 and 1, ensuring no inherent similarity between vectors and difference between  
119 classes (see **Materials and Methods** for details). Expectedly, the observed chance level of  
120 PRED, PC, and COS was nearly 0 for both the 2-class and 5-class datasets; it was greater  
121 than 0 for MAN, EUC, and CHEB for both types of datasets (**Figure 1b**). Moreover, MAN,  
122 EUC, and CHEB's chance levels were different for the datasets with different numbers of  
123 classes (**Figure 1b**). This difference occurs because the distances between vectors depend on  
124 the vectors' sizes; we can more directly observe this change in chance levels with  
125 untransformed MAN, EUC, and CHEB metrics, all of which showed larger values with more  
126 classes (**Supplementary Figure 1b**). We tried to normalize these metrics according to the  
127 number of classes – for example, by dividing MAN by the number of classes or dividing  
128 EUC values by the square root of the number of classes. Although these normalizations  
129 reduced the overall differences between the chance levels for different numbers of classes,  
130 the differences remained significant (**Supplementary Figure 1c**). Thus, distance-based  
131 metrics do not provide a stable chance level.

132 Another important consideration for assessing a metric's utility is its ability to report the level  
133 of similarity for a dataset, and its modifications, in a way that matches intuition. We had  
134 previously reported PRED's advantages over PC in calculating stereotypy (Mittal et al.,  
135 2020). Here, we extend this analysis to include the other metrics. If the responses in a vector  
136 are the same for both classes, PRED reports a value of 0; however, PC is undefined, and COS  
137 reports a high value (**Supplementary Figure 1a (ii)**). If the two vectors exhibit opposite  
138 patterns across the classes (**Supplementary Figure 1a (iii)**), PRED and PC appropriately  
139 quantify the similarity as -1. COS, however, still reports a value close to 1, which does not  
140 match the intuitive difference between the two vectors. The distance-based metrics also fail to  
141 capture this difference: they report the same values of similarity in **Supplementary Figure**  
142 **1a (iii)** and **(iv)**, even though in one case the vectors exhibit opposite patterns and in the other  
143 case they exhibit similar patterns across the two classes. If we linearly transform all the  
144 values in a dataset in the same manner, intuitively, the similarity between them should not  
145 change. Except for COS, all metrics are stable to global translational change, i.e., the addition  
146 of a constant to all the values in the dataset (**Supplementary Figure 1a (v)** compared to **(iv)**).  
147 Similarly, all metrics, except MAN, EUC, and CHEB, are stable to scaling modifications, i.e.,  
148 multiplication of the entire dataset by a constant value (**Supplementary Figure 1a (vi)**  
149 compared to **(iv)**).

150 Overall, PRED behaved intuitively for various modifications within the datasets, while each  
151 of the other metrics deviates from the intuition in one or more cases (summarized in **Table**  
152 **1**). As the distance-based metrics (MAN, EUC, and CHEB) lack a stable chance level, are not  
153 sensitive to patterns in the dataset, and are not robust to simple scaling transformations, we  
154 exclude them from further consideration as metrics of similarity.

155 Experimental datasets are often noisy. With any metric, we expect the similarity between two  
156 vectors to decrease as the noise level in the dataset increases, eventually reaching the chance  
157 level for extreme levels of noise. We studied how PRED, PC, and COS behaved for different  
158 noise levels using two parameters: dynamic range and variability. Here, dynamic range  
159 denotes the range of noise levels within which a metric exhibits unsaturated values and thus  
160 remains useful. Variability represents the sensitivity of a metric to noisy data: we consider a  
161 metric to have high variability if it shows very different values for different samples of the  
162 data at a given noise level. We quantified variability as the percent standard deviation over  
163 repeated simulations with noise at the mid-point of the dynamic range (see **Materials and**  
164 **Methods**). A useful metric should have a high dynamic range and low variability. We  
165 measured both these parameters for PRED, PC, and COS in a simulated dataset (see  
166 **Materials and Methods**) with increasing noise levels (**Figures 1c—e**). We found that PRED  
167 exhibited the highest dynamic range and lowest variability among all the metrics (**Figure 1f**).  
168 Even for simulated datasets with different base means, PRED was consistently more robust  
169 than the other metrics (**Supplementary Figures 1d, e**). Thus, PRED remains informative  
170 across a relatively large range of noise levels in the dataset and provides a relatively stable  
171 estimate of similarity.



## 172 **PRED for behavioral similarity assessment**

173 We previously applied PRED to comparing the similarity of neural response patterns to an  
174 odor set across individuals (Mittal et al., 2020). However, in principle, it can be applied to  
175 any dataset where the data are arranged as vectors (each vector's length equals the number of  
176 classes). Many behavioral studies examine if the behavioral outcomes of multiple individuals  
177 are similar over different time points. Here, one could consider the individuals as classes and  
178 each time point as a vector. Honegger et al. (Honegger et al., 2020) measured the preference  
179 indices of 141 *Drosophila* flies in a two-choice assay between two odors (3-octanol versus 4-  
180 methylcyclohexanol) over two different time points 24-hours apart (**Figure 2a**). They used  
181 PC to compare the similarity of preference index vectors across the two time points and  
182 found a moderate positive value of 0.35 (Honegger et al., 2020). Using PRED on the same  
183 data, we observed a value of 0.19, indicating a moderate similarity between behavioral  
184 preferences across the two time points.

185 Our results above (**Figure 1f**) have indicated that PRED is more stable than PC for noisy  
186 data. Therefore, we reasoned that it would also be more robust when working with  
187 incomplete datasets. The 141-fly behavioral dataset provided a suitable test case for this idea.  
188 We randomly selected 70 flies from the dataset and calculated the similarity of the preference  
189 index vectors at the two time points using PRED and PC. This random sampling was repeated  
190 20 times, each resulting in a different value of PRED and PC. Even with incomplete datasets,  
191 both metrics reported significant similarity:  $0.20 \pm 0.04$  ( $P = 8.9 \times 10^{-15}$ ,  $n = 20$ ; one sample  
192 t-test compared to 0) for PRED; and  $0.37 \pm 0.10$  ( $P = 6.8 \times 10^{-13}$ ,  $n = 20$ ) for PC. Note that  
193 the PRED values were less variable (smaller s.d.) over the repeated samplings. Even the  
194 coefficient of variation, defined as  $COV = \frac{s.d.}{mean}$ , over these 20 samplings was smaller for  
195 PRED (0.21) than PC (0.27) (**Figure 2b**). Since these observed values of the COV may  
196 depend on the specific 20 samplings that occurred, we repeated the whole process of 20  
197 samplings a total of 50 times and each time calculated the COVs for both metrics. This  
198 analysis confirmed that the COV was consistently lower for PRED ( $P = 2.8 \times 10^{-15}$ ,  $n = 50$ ,  
199 two-sample paired t-test; **Figure 2c**). Thus, PRED provides a relatively stable estimate of  
200 similarity for partial samplings of the dataset.

201

## 202 **Similarity in multi-dimensional data**

203 So far, we have calculated similarity between two vectors where each vector contains a set of  
204 values corresponding to the set of classes—for example, comparing the response of a neuron  
205 to 2 stimuli (classes) in 2 individuals (vectors). This formatting is feasible for datasets where  
206 the response is a single number, such as the total number of spikes (or the net firing rate)  
207 evoked by a stimulus within a pre-defined time window. However, one may want to look at  
208 the response in finer detail, for example, by considering the temporal pattern of spikes evoked  
209 by the stimulus. We can represent the temporal pattern as a set of numbers by dividing the  
210 time window into, say, 10 bins and then counting the spikes in each bin. Thus, the response to

211 a stimulus is now itself a 10-element vector rather than a single number (**Figure 3a**). In this  
212 case, if we want to compare the responses to a set of stimuli in two individuals, we need to  
213 compare two vectors of vectors rather than two vectors of numbers (**Figure 3a**).

214 Although correlation is frequently used to quantify the similarity between vectors, it is not  
215 equipped to handle vectors of vectors. A common modification to use correlation in such  
216 cases is concatenating the internal vectors within the outer vector to result in a single (and  
217 long) vector. In the example discussed earlier, it would mean combining the two 10-element  
218 vectors corresponding to the two stimuli to obtain a 20-element vector for each individual and  
219 then calculating the correlation between the 20-element vectors of the two individuals  
220 (**Figure 3a**). On the other hand, PRED is natively equipped to handle vectors of vectors and  
221 does not require concatenation: it involves calculating Euclidean distances between the  
222 values, which we can do irrespective of whether the values are single numbers or vectors. In  
223 the example discussed above, we can calculate  $D_1$  and  $D_2$  for PRED based on the 10-  
224 dimensional Euclidean distances between the binned responses and then PRED using the  
225 regular formula,  $\frac{D_2 - D_1}{D_2 + D_1}$  (**Figure 3a**).

226 We used both PRED and PC to compare the firing rates or the 10-bin temporal patterns  
227 evoked by odors in different individuals (see **Materials and Methods**). We performed this  
228 analysis in two different datasets: the olfactory response of mushroom body output neuron,  
229 bLN1, in locusts (Gupta and Stopfer, 2014) and four different projection neurons in  
230 *Drosophila* (Shimizu and Stopfer, 2017). We used a 2-second window after odor-onset to  
231 calculate the responses; in these datasets, the responses typically returned to baseline within 2  
232 seconds in response to the 1-s odor pulse. Therefore, we can consider any spikes observed  
233 after this window as a part of the background spiking. For the temporal response, we divided  
234 this response into ten bins, each of length 200-ms (**Figure 3a**). Both PRED and PC revealed  
235 significant similarities between individuals and showed that the similarity was slightly lower  
236 when considering the temporal patterns instead of only the firing rates (**Figure 3b** and  
237 **Supplementary Figures 2a—d**).

238 Although PRED and PC behaved similarly in this analysis, PC can run into problems because  
239 of the concatenation step. Concatenation removes the distinction between the values  
240 belonging to different bins within the same class and the values belonging to different  
241 classes. For example, after concatenation, analyzing the 10-element temporal responses to 2  
242 stimuli becomes identical to analyzing the firing rate responses to 20 independent stimuli,  
243 with each element contributing equally to the correlation. To illustrate why this can be  
244 problematic, we consider the case when the temporal response includes bins beyond the  
245 stimulus-evoked response; these bins would be mostly empty except for some noise. Since  
246 empty bins are similar by nature, including such bins in the response vectors and effectively  
247 treating them as independent stimuli after concatenation would spuriously increase the  
248 observed correlation.

249 In contrast, the calculation of Euclidean distances in PRED would be minimally affected by  
250 the empty bins: the distances would only become slightly noisier by the noise in the empty

251 bins. Thus, PRED would report slightly lower similarity, which is a more intuitive outcome  
252 given the inclusion of irrelevant bins. To test these predictions in the actual datasets analyzed  
253 here, we included extra bins after the initial 10 bins of 200 ms duration. For example, in an  
254 11-bin response, the first 10 bins would contain the first 2-s response after odor onset, while  
255 the last bin would contain an extra 200-ms response from 2 to 2.2-s after odor onset. Since  
256 the stimulus-evoked response typically lasted for less than 2 s, the extra bins included after  
257 the 2-s response are usually empty except for some noise. We found that, as predicted, the PC  
258 values increased as we added more and more extra bins in the response, whereas the PRED  
259 values decreased (**Figure 3c** and **Supplementary Figures 2f—i**). We further simulated a  
260 dataset containing two odors and ten individuals. The first 10 bins contained a simulated  
261 temporal response, and the subsequent bins contained random noise (see **Materials and**  
262 **Methods**). There was a noticeable increase in the PC values in these simulations with an  
263 increasing number of extra bins (**Figure 3d**). The effect became more pronounced when we  
264 added empty bins (i.e., bins with a value of 0) instead of bins with normally distributed noise.  
265 In this case, PRED values were constant as the empty bins did not affect the distances in  
266 PRED calculations (**Supplementary Figure 2e**). These results illustrate the pitfalls in using  
267 concatenated vectors in PC and suggest that PRED is a better alternative when working with  
268 multi-dimensional data.

269 Another type of multi-dimensional data is population-level data, i.e., the response of, say, 6  
270 neurons from the same neural layer from two individuals responding to two stimuli. To  
271 analyze such a case, we can either calculate the similarity separately for each neuron and then  
272 take the average or directly consider the 6-element population response vector for each  
273 individual and odor. We used PRED to compare these two approaches, using a published  
274 dataset of calcium imaging responses of 37 antennal lobe glomeruli responding to 36 pure  
275 odors in 61 individuals (Badel et al., 2016). The similarity observed between individuals  
276 using the population vectors was significantly more than the average similarity of neurons  
277 considered separately (0.37 compared to  $0.25 \pm 0.10$ ,  $P = 1.7 \times 10^{-10}$ ,  $n = 37$ ; one-sample t-  
278 test; **Figure 3e**). These results suggest that the combined cell population preserves more  
279 similarity within the system than individual cells, echoing previous studies' results (Mittal et  
280 al., 2020). The results also illustrate the usefulness of PRED in analyzing population-level  
281 data.

282

## 283 **Class separability**

284 The datasets we have considered so far had a class-vector structure (as shown in **Figure 1a**):  
285 multiple vectors (rows), each containing values for multiple classes (columns). The value of  
286 PRED for such a dataset depends on, and thus tells us about, both the similarity between the  
287 vectors and the separability of the classes. (Contrast this with Euclidean distance, which tells  
288 us only about the similarity between the vectors but is a poor indicator of class separability,  
289 as can be seen by comparing **Supplementary Figure 1a (iii)** and **(vi)**). In these datasets,  
290 there is a correspondence between the  $i^{th}$  value in class 1 and the  $i^{th}$  value in class 2, as they



291 both belong to the same vector in row  $i$  (which could be an individual, a time-point, or any  
292 other variable depending on the experimental context). However, many datasets do not have  
293 this correspondence (i.e., there are no row-vectors) — for example, in neuroscience, one  
294 often measures the responses of a neuron or a brain region to different stimuli (classes) and  
295 takes multiple measurements (called trials or samples) for each stimulus. In such cases we are  
296 left with only classes (columns), with each class containing multiple values (as shown in  
297 **Figure 4a**). This formatting is commonly used in datasets with repeat measurements over  
298 multiple classes. Here, the numbers of samples for different classes do not have to be  
299 identical. Each sample value within a class may be a single number (e.g., the firing rate of a  
300 neuron or the preference index of an animal) or a set of numbers (e.g., a binned temporal  
301 response or a population response). Assuming that the samples within a class are generated  
302 under identical experimental conditions and that the samples in different classes are generated  
303 independently, there is no logical correspondence between the  $i^{th}$  sample in class 1 and the  
304  $i^{th}$  sample in class 2. We will refer to such datasets as class-sample datasets. In such datasets,  
305 one often wants to know about the separability of the classes.

306 A similar requirement arises when evaluating the output of unsupervised clustering  
307 algorithms, which use statistical methods to divide a collection of values into different  
308 clusters. The resulting clusters are analogous to classes in the above formulation, and their  
309 assigned members are analogous to samples. Here also, one often wants to know how well  
310 separated the observed clusters are. For example, Karagiannis et al. classified neuropeptide  
311 Y-expressing neocortical interneurons into 3 different types based on their morphology using  
312 a K-means clustering algorithm (Karagiannis et al., 2009). They then used the Silhouette  
313 index (Rousseeuw, 1987) to evaluate the quality of the clustering obtained. Another study  
314 used the Silhouette index to assess the efficiency of single nucleotide polymorphism  
315 genotyping assays in dividing samples into 3 different groups: homozygous for the first  
316 allele, homozygous for the second allele, or heterozygous (Lovmar et al., 2005). Apart from  
317 the Silhouette index (Rousseeuw, 1987), an evaluation of a clustering technique's efficacy  
318 can be made using other internal clustering validation indices like the Davies-Bouldin index  
319 (Davies and Bouldin, 1979) or the Dunn's index (Dunn, 1974). Another method commonly  
320 used to measure class separability is Euclidean template matching (ETM), which involves  
321 classifying each value based on its Euclidean distance from class templates (constructed from  
322 the remaining data) and then calculating the average accuracy from these classifications  
323 (Stopfer et al., 2003).

324 Since the PRED value for a class-vector dataset depends on class separability, we asked  
325 whether PRED can also be used as a measure of class separability in class-sample datasets  
326 (**Figure 4a**). We compared PRED to five commonly used metrics: Silhouette index (SIL),  
327 Davies-Bouldin index (DBI), Dunn's index (DUNN), ETM, and Calinski-Harabasz index  
328 (CH) (see **Materials and Methods** for a description of each metric). As an initial test of  
329 PRED's feasibility for this application, we used two different datasets containing repeated  
330 responses to different odors. We obtained one dataset from the identified *bLNI* neuron in  
331 locusts (Gupta and Stopfer, 2014) and another from four identified projection neurons in  
332 *Drosophila* (Shimizu and Stopfer, 2017). Each dataset contains the response from multiple

333 individuals; we compared the odor separability calculated using PRED and the other metrics  
334 for each individual. We found that PRED values were somewhat correlated with the values  
335 from other metrics in both the datasets (**Figures 4b—f** and **Supplementary Figures 3a—e**).  
336 (Note that the correlation with DBI is negative because a lower DBI value indicates a higher  
337 separability, whereas the opposite is true for PRED and the other four metrics). These  
338 correlations with the established metrics suggested that PRED might also be useful as a  
339 metric of class separability. To explore this further, we compared PRED's performance with  
340 the other metrics in various situations.

341 As discussed in the analysis of class-vector datasets, a key feature of any metric is its chance  
342 level. For evaluating the chance level of separability metrics in class-sample datasets, we  
343 simulated datasets containing clusters (classes) of points with fixed radii on a 2-d plane and  
344 different levels of noise (**Supplementary Figure 3f**; see **Materials and Methods** for  
345 details). As we increase the noise in the simulated dataset, the classes lose their separability  
346 (**Supplementary Figures 3f—h**). We used datasets with extremely high noise levels to  
347 calculate the chance level of each of the six metrics. Further, we checked how the chance  
348 levels depend on the number of classes in the dataset. PRED showed a chance level close to  
349 0, regardless of the number of classes. CH showed a chance level greater than 0 that was not  
350 different for 2-class or 5-class datasets (**Figures 4g, I**). However, the chance levels of the  
351 other four metrics changed significantly with the number of classes (**Figures 4h—k**).

352 Imagine a large dataset containing many classes where any two classes have the same level of  
353 separability, whose value is not known to us. Further, imagine that, for practical reasons, we  
354 have access to only a subset of the dataset covering some of the classes, and our task is to use  
355 different metrics to estimate the class separability. An ideal metric should estimate the same  
356 underlying class separability, regardless of the number of classes available in our subset. To  
357 check how the six metrics under consideration perform on this criterion, we simulated a  
358 dataset with a low level of noise (thus with reasonable class separability) and varied the  
359 number of classes. We found that the separability reported by all metrics except PRED varied  
360 with the number of classes (**Figure 5a**).

361 CH values decreased with the number of classes when we had 2 samples per class but not  
362 when we had 10 samples per class (**Figure 5a**; **Figure 4I** also had 10 samples per class,  
363 which explains no change in the CH chance level). This result indicated that the number of  
364 samples could also bias the value of a metric. Ideally, the separability of the classes should  
365 not depend on how many samples are available for each class. For example, our estimate of  
366 how well a neuron can differentiate two sensory stimuli (a property of the neuron and the  
367 stimuli) should not be biased by the number of recording trials available (an experimental  
368 parameter). We performed another set of simulations with 2 classes and an increasing number  
369 of samples per class. We found that CH, ETM, and DUNN values varied significantly with  
370 the number of samples (**Figure 5b**), while PRED, SIL, and DBI were relatively stable. We  
371 conclude that PRED provides an unbiased estimate of class separability regardless of the  
372 number of classes or the number of samples per class. Therefore, we can reliably use it with  
373 datasets of all sizes.

374 We next studied the stability of each metric against noisy data by checking the dynamic range  
375 and the variability at the midpoint of the dynamic range. We simulated datasets with noise  
376 levels ranging from zero (highly separable classes) to very high (poorly separable classes). As  
377 before, we estimated the dynamic range as the range of noise levels for which a metric  
378 remained unsaturated and variability as the percent standard deviation over repeated  
379 simulations with noise at the mid-point of the dynamic range (**Figures 6a—f**). PRED and SIL  
380 showed the best combination of large dynamic range and small variability (**Figure 6g**).  
381 DUNN had the lowest dynamic range and high variability, while DBI exhibited a high  
382 dynamic range but also the highest variability (**Figure 6g**). We used the *Drosophila* and  
383 locust datasets to complement the simulation results. We added increasing amounts of noise  
384 to each value in the datasets and then compared the metrics (**Supplementary Figure 4**; see  
385 **Materials and Methods**). Again, PRED and SIL exhibited large dynamic ranges and small  
386 variabilities in all cases. DUNN and DBI showed a high dynamic range in some cases but  
387 were the worst performers in variability in most neurons. Overall, PRED and SIL appear to  
388 be the most robust metrics in handling noisy datasets. Considering that SIL values (including  
389 the chance level) depend on the number of classes, as discussed above, PRED appears to be  
390 the best among the considered metrics for quantifying class separability (summarized in  
391 **Table 2**).

392 Class separability depends on how different the values are across the classes and how similar  
393 they are for different samples within each class. PRED, thus, may be a useful metric when  
394 both within-class similarity and across-class differences are analyzed simultaneously.  
395 Kermen et al. (Kermen et al., 2020) looked at zebrafish olfactory behaviors elicited by a set  
396 of 18 odors in different individuals while performing 4 repeated trials with each odor. They  
397 calculated the intra-individual similarity by correlating the behavioral responses across all  
398 pairs of trials for each individual and the inter-individual similarity by correlating the trial  
399 averaged response of all pairs of individuals. Then they looked at pairs of these two similarity  
400 values to examine how consistent the responses produced by each odor were within and  
401 across individuals. If one wants to know which odors produce relatively similar responses  
402 within individuals but different across individuals, PRED can provide the answer with a  
403 single number. We calculated PRED considering individuals as classes and trials as samples  
404 (**Figure 7a**; see **Materials and Methods**). We found that the behavioral responses were  
405 relatively different across individuals and consistent across trials for these odors: cadaverine  
406 ( $0.39 \pm 0.34$ ,  $P = 1.8 \times 10^{-6}$ ,  $n = 28$ ), blood ( $0.39 \pm 0.35$ ,  $P = 8.2 \times 10^{-4}$ ,  $n = 15$ ), skin  
407 ( $0.26 \pm 0.32$ ,  $P = 0.007$ ,  $n = 15$ ), bile ( $0.17 \pm 0.26$ ,  $P = 4.2 \times 10^{-4}$ ,  $n = 36$ ), sperm ( $0.15 \pm 0.35$ ,  $P$   
408  $= 0.007$ ,  $n = 45$ ), cysteine ( $0.13 \pm 0.38$ ,  $P = 0.05$ ,  $n = 36$ ), and arginine ( $0.12 \pm 0.26$ ,  $P = 0.01$ ,  $n$   
409  $= 36$ ) (**Figure 7b**).

410

### 411 **Using PRED for assessing individuality**

412 Honegger et al. (Honegger et al., 2020) observed that odor preferences of *Drosophila* varied  
413 more across individuals than across trials within an individual. Consistent with this, they also

414 found that the odor responses of the projection neurons were also more variable across  
415 individuals than across trials, suggesting that this response individuality may underlie the  
416 behavioral individuality. The behavioral individuality depended on serotonin: it reduced  
417 when the flies were fed alpha-methyl tryptophan, a serotonin synthesis blocker. However,  
418 somewhat unexpectedly, they did not detect a reduction in the response individuality in the  
419 presence of the serotonin blocker. Their analysis used principal component analysis and  
420 Bayesian modeling to compute inter-fly and intra-fly distances. Since quantifying  
421 individuality requires an assessment of inter-individual differences relative to intra-individual  
422 differences, we reasoned that individuality could be aptly described by class separability,  
423 where the individuals are classes, and the trials are samples within each class. We reanalyzed  
424 their data using individual-trial (class-sample) PRED to quantify the individuality of the PN  
425 responses to different odors (**Figure 7c**; see **Materials and Methods**). In the wild-type flies,  
426 we observed that 50% (84 out of 168) of the PN-odor responses were significantly separable  
427 across individuals (**Figure 7d**), matching the conclusions of Honegger et al. However, in  
428 serotonin-blocked flies, this fraction reduced to only ~24% (40 out of 168; **Figure 7e**) even  
429 though the original analysis was not able to uncover this reduction. Thus, our reanalysis of  
430 response individuality shows that serotonin indeed affects the PN response individuality. By  
431 resolving the contradiction between the behavioral data and the PN response data in the  
432 presence of serotonin blockage, our analysis using PRED lends additional support to the idea  
433 of Honegger et al. (Honegger et al., 2020) that PN response individuality determines  
434 behavioral individuality.

435

## 436 **Using PRED for analyzing connectomic data**

437 Recent advances in high-throughput electron microscopy and image segmentation methods  
438 have made it possible to reconstruct neuronal morphologies and connections in large brain  
439 areas. For *Drosophila*, two public datasets, namely the full adult fly brain or FAFB (Zheng et  
440 al., 2018) and the Hemibrain (Scheffer et al., 2020), have recently become available. As these  
441 datasets are generated from two different individuals, they provide an opportunity for  
442 measuring stereotypy in the connectivity patterns of neurons across individuals. A recent  
443 study by Schlegel et al. (Schlegel et al., 2021) used these two datasets to measure stereotypy  
444 in the input connections received by the lateral horn neurons (LHNs) from the projection  
445 neurons (PNs). For each LHN, they calculated a vector of connectivity with different types of  
446 PNs and used the cosine metric (COS) to estimate the similarity between such vectors. They  
447 demonstrated stereotypy in the inputs of LHNs by a combination of two results: (i) when  
448 comparing LHNs belonging to the same cell type, the COS values for LHNs across the two  
449 datasets were high and similar to the COS values for LHNs within a dataset; and (ii) when  
450 comparing LHNs belonging to different cell-types, the COS values for LHNs across the two  
451 datasets were low and similar to the COS values for LHNs within a dataset.

452 PRED allows one-shot quantification of stereotypy in this case with a single number. Based  
453 on their morphologies and connections to other neurons, the LHNs have been grouped into



454 ‘connectivity types,’ which are further grouped into ‘regions,’ ‘tracts,’ and ‘cell types’ in the  
455 increasing order of hierarchy (see **Materials and Methods**). Although it has not been  
456 possible to match the neurons in the two datasets unambiguously, these higher-order  
457 groupings have been labeled in both datasets. We computed a 57-length glomerular input  
458 vector for each group by averaging the connectivity vectors of all LHNs belonging to the  
459 group (**Figure 8a**). To estimate stereotypy in the glomerular input vectors of groups at a  
460 particular hierarchy level, we calculated the group-dataset (class-vector) PRED (**Figure 8a**).  
461 At the level of ‘connectivity types,’ we found that the PRED value was  $0.56 \pm 0.25$  ( $P =$   
462  $4.4 \times 10^{-193}$ ,  $n = 496$ ), notably higher than the chance level of 0, suggesting that the  
463 averaged connectivity vectors were separable across connectivity types and similar across the  
464 two datasets. Similarly, high PRED values were also seen at other grouping levels (cell type:  
465  $0.56 \pm 0.25$ ,  $P = 1.4 \times 10^{-147}$ ,  $n = 378$ ; tract:  $0.61 \pm 0.16$ ,  $P = 1.4 \times 10^{-40}$ ,  $n = 66$ ; region:  
466  $0.56 \pm 0.18$ ,  $P = 5.5 \times 10^{-4}$ ,  $n = 6$ ), confirming the stereotypy in the connectivity patterns of  
467 LHNs groups across the two databases.

468 The above analysis compared the averaged glomerular connectivity patterns of different  
469 groups. Next, we sought to assess whether the glomerular connectivity patterns of different  
470 neurons within a group were more consistent than the patterns of neurons across different  
471 groups at the same hierarchy level. This could be easily quantified as group-separability using  
472 group-neuron (class-sample) PRED. In both the datasets, we found that the ‘connectivity  
473 types’ were highly separable (FAFB: PRED =  $0.47 \pm 0.21$ ,  $P = 2.1 \times 10^{-15}$ ,  $n = 36$ ;  
474 Hemibrain: PRED =  $0.50 \pm 0.25$ ,  $P = 3.7 \times 10^{-96}$ ,  $n = 276$ ; **Figure 8b**). Similarly, the cell  
475 types were also highly separable (FAFB: PRED =  $0.44 \pm 0.19$ ,  $P = 8 \times 10^{-10}$ ,  $n = 21$ ;  
476 Hemibrain: PRED =  $0.50 \pm 0.21$ ,  $P = 6.5 \times 10^{-88}$ ,  $n = 210$ ). The separability reduced as we  
477 went to higher levels in the group hierarchy, namely the ‘tracts’ (FAFB: PRED =  $0.11 \pm 0.15$ ,  
478  $P = 8.4 \times 10^{-5}$ ,  $n = 36$ ; Hemibrain: PRED =  $0.18 \pm 0.14$ ,  $P = 4.3 \times 10^{-11}$ ,  $n = 45$ ) and the  
479 ‘regions’ (FAFB: PRED =  $0.05 \pm 0.06$ ,  $P = 0.081$ ,  $n = 6$ ; Hemibrain: PRED =  $0.06 \pm 0.3$ ,  $P =$   
480  $0.0049$ ,  $n = 6$ ). This reduction in class separability reflects the increasing diversity of neurons  
481 within the higher-level groups. Overall, these results demonstrate how class-sample PRED  
482 can be used as a sensitive and easy-to-use metric of class separability.

483

## 484 **Discussion**

485 Overall, we found that Pairwise Relative Distance (PRED) is a robust metric for quantifying  
486 vector similarity and class separability in class-vector datasets and offers several advantages  
487 over distance-based metrics, Pearson’s correlation, or cosine similarity. Importantly, PRED  
488 quantified the similarity in a consistent way close to our intuitive understanding of the data.  
489 Datasets in different studies often vary in terms of their size and the scale of the responses. If  
490 the similarity metric is affected by these parameters, it becomes difficult to compare the  
491 results obtained across studies. PRED, however, remained agnostic to the size of the dataset  
492 and was unchanged with global modifications of the data (**Figure 1** and **Supplementary**  
493 **Figure 1**). We can, thus, directly compare PRED values obtained from different studies.



494 Experimental studies may be limited in the amount of data that they can collect; in terms of,  
495 for example, how many different stimuli one can present, or how many individuals can study,  
496 or how many trials one could perform, and so on. Also, experimental data is subject to noise  
497 from multiple sources. Thus, it is desirable to analyze datasets with a metric that is robust to  
498 noise. In our study, PRED exhibited the largest dynamic range and the lowest variability  
499 among the metrics tested. It also worked well with incomplete datasets (**Figures 1, 2, and**  
500 **Supplementary Figure 1**).

501 Many metrics are available for calculating the similarity of vectors when each value within  
502 the vector is a scalar quantity (a number). However, we cannot directly use these metrics  
503 when each value within the vector is itself a vector (a set of numbers), as is the case with  
504 temporally patterned neural responses or population responses. One could forcibly convert  
505 the vector of vectors into a long vector of numbers through concatenation. However,  
506 concatenated vectors lose the distinction between classes and the elements of values within a  
507 class. As we showed by simulating increasingly longer temporal patterns, this can lead to an  
508 inaccurate estimation of similarity. On the other hand, PRED provides a more straightforward  
509 and intuitive method for analyzing multi-dimensional data while preserving the inherent  
510 relations between different dimensions (**Figure 3 and Supplementary Figure 2**).

511 We found that PRED also works well for analyzing class separability in class-sample  
512 datasets, as the results with PRED were well correlated with those obtained from other  
513 commonly used metrics. PRED provided a stable chance level and was unaffected by the  
514 dataset's size, whereas most of the other metrics that we tested varied with an increase in the  
515 number of classes or samples. We tested the robustness of several internal clustering  
516 validation metrics to noisy datasets. In these analyses using simulated and experimental data,  
517 PRED was consistently among the metrics with the highest dynamic range and the lowest  
518 variability. Thus, PRED presents a consistent and more reliable alternative for evaluating  
519 class separability in class-sample datasets (**Figures 4 – 8 and Supplementary Figures 3—**  
520 **5**).

521 When dealing with large datasets, one consideration in choosing a metric is its computational  
522 time complexity. Since PRED calculates the similarity iteratively for all combinations of  
523 pairs of classes and pairs of vectors, its time complexity is of the order of  $O\left(\binom{m}{2} \times \binom{n}{2}\right) =$   
524  $O(m^2n^2)$ , where  $m$  and  $n$  are the numbers of classes and vectors, respectively. Thus, the  
525 time required to compute PRED increases polynomially with an increase in the dataset's size.  
526 Other class-vector metrics including Pearson's correlation, cosine similarity, and distance-  
527 based metrics have  $O(mn^2)$  time complexity. However, datasets in many applications are  
528 small enough ( $m, n \leq 100$ ) that the time complexity of PRED would not become a limiting  
529 consideration.

530 We originally designed PRED for class-vector datasets, in which there is a correspondence  
531 between the  $i^{th}$  element in class 1 and the  $i^{th}$  element in class 2, as both elements belong to  
532 the same vector (row). PRED calculation makes use of this correspondence when making the  
533 2x2 matrices for a pair of classes: if a 2x2 matrix has the  $i^{th}$  and the  $j^{th}$  values from class 1,

534 it must have the  $i^{th}$  and the  $j^{th}$  values from class 2). In class-sample datasets, this  
535 correspondence across classes is absent, as there is no ordering among the class elements – all  
536 samples are random replicates. This lack of order poses a dilemma while calculating PRED:  
537 which pair of values in class 2 should we use for making the 2x2 matrix with a particular pair  
538 of values in class 1? We overcome this dilemma by considering all possible pairs from class 2  
539 iteratively for a given pair of values in class 1. This method (‘exhaustive PRED’) increases  
540 the time complexity from  $O\left(\binom{m}{2} \times \binom{n}{2}\right)$  to  $O\left(\binom{m}{2} \times \binom{n}{2}^2\right)$  for class-sample datasets,  
541 assuming each of the  $m$  classes has  $O(n)$  elements (**Supplementary Figure 5a**). In practice,  
542 the extra time required for ‘exhaustive PRED’ would be noticeable only for large datasets  
543 with hundreds of classes and samples. The calculation can be made faster using an  
544 approximation (‘fast PRED’). In ‘fast PRED,’ we assign an arbitrary order to the elements in  
545 each class (e.g., the order in which the values were saved) and then create 2x2 matrices in the  
546 same way as is done in class-vector datasets: when we take the  $i^{th}$  and the  $j^{th}$  values from  
547 class 1, we also take the  $i^{th}$  and the  $j^{th}$  values from class 2. Using simulations (see **Materials**  
548 **and Methods**), we found that the difference between the ‘exhaustive PRED’ and the ‘fast  
549 PRED’ values was ~3% for datasets with more than 15 samples (**Supplementary Figure 5b**).  
550 Changing the ordering of elements within classes did not have a noticeable effect on the  
551 value of PRED. Thus, we can efficiently and reliably compute PRED for large class-sample  
552 datasets.

553 Class-sample PRED essentially compares the within and across class variation of samples. As  
554 classification is a very commonly used operation, there has been a strong interest in  
555 comparing various metrics under different scenarios (Arbelaitz et al., 2013; Brun et al., 2007;  
556 Guerra et al., 2012; Gurrutxaga et al., 2011; Niemelä et al., 2018). Apart from the metrics that  
557 we have already compared with PRED, other metrics with similar approaches, like the t-  
558 statistic or Fisher discriminant, can potentially be used for analyzing class-sample datasets.  
559 However, these metrics have their drawbacks. The calculation and the interpretation of the t-  
560 statistic depend on the degree of freedom, which is a function of the number of samples  
561 observed. The discriminant analysis assumes a linear separation between the classes and thus  
562 might not be ideal for neural datasets. Another approach, formulated by Huerta et al. (Huerta  
563 et al., 2004), also quantifies intra-class and inter-class differences. They calculated average  
564 within-class ( $D_{intra}$ ) and across-class ( $D_{inter}$ ) distances, similar to our  $D_1$  and  $D_2$   
565 calculations. They then quantified the similarity across classes by measuring  $D_{inter} - D_{intra}$   
566 normalized by the maximum expected value of this difference. The normalization procedure  
567 is highly dependent on the type of system under consideration, and it might not be possible to  
568 calculate the denominator in many cases. PRED is self-normalizing and system agnostic,  
569 providing a consistent estimate of class separability for any dataset.

570 So far, we have computed  $D_1$  and  $D_2$  as the Euclidean distances between within-class and  
571 across-class values. In principle, one can use any distance measure in place of Euclidean  
572 distances for calculating PRED. For example, one can use Mahalanobis distance to account  
573 for different variabilities of the various dimensions of a response or Hamming distance to  
574 compare datasets with binary or categorical values. For temporal data, instead of binning the

575 responses, one could use methods like the Victor-Purpura (Victor and Purpura, 1997, 1996)  
576 or the van Rossum (Rossum, 2001) distances to calculate the distance between spike trains.  
577 This flexibility in the choice of the distance metric may help in the future in optimizing  
578 PRED for different use cases.

579

## 580 **Materials and Methods**

### 581 **Class-vector PRED**

582 We generalized the definition of PRED from our previous work (Mittal et al., 2020) to all  
583 class-vector datasets. We considered all possible combinations of pairs of vectors and pairs of  
584 classes to calculate the PRED value. For each  $2 \times 2$  matrix thus obtained, we computed two  
585 distances (**Figure 1a**):  $D_1 = (A1 - B1)^2 + (A2 - B2)^2$  is the sum of the squared Euclidean  
586 distances between the values to the same classes in different vectors;  $D_2 = (A1 - B2)^2 +$   
587  $(A2 - B1)^2$  is the sum of the squared distances between the values belonging to different  
588 classes in different vectors. We used the ratio  $\frac{D_2 - D_1}{D_2 + D_1}$  to estimate the PRED value in each  $2 \times 2$   
589 matrix. To obtain the final PRED value for a particular dataset, we first averaged the values  
590 over all class pairs before averaging over all vector pairs. Cases with missing data were  
591 ignored for the calculation of the mean. Note that in the calculations described here, the  
592 Euclidean distances can be easily calculated even if the values ( $A1, B1, A2, B2$ ) are not  
593 numbers but are equal-sized vectors (see **Figure 3a** for an example). PRED ranges between 1  
594 and -1, where 1 indicates that the vectors have identical values and patterns across classes, 0  
595 indicates that the vectors have no similarity and have random patterns across the classes, and  
596 -1 indicates that the vectors have exactly opposite patterns across the classes.

### 597 **Class-sample PRED**

598 We used a slightly modified method of calculating PRED (labeled ‘exhaustive PRED’) for  
599 class-sample datasets (**Figure 4a**). The calculation of  $D_1$  and  $D_2$  and the ratio  $\frac{D_2 - D_1}{D_2 + D_1}$  remained  
600 unchanged. The difference here lay in the creation of  $2 \times 2$  matrices: for each pair of classes,  
601 any two samples (say, 1A and 1B) in class  $i$  could be combined with any two samples (say,  
602 2A and 2B) in class  $j$ , to create two possible matrices,  $\begin{bmatrix} 1A & 2A \\ 1B & 2B \end{bmatrix}$  or  $\begin{bmatrix} 1A & 2B \\ 1B & 2A \end{bmatrix}$ . This results  
603 in a total of  $\binom{n_i}{2} \cdot \binom{n_j}{2} \cdot 2$  matrices for classes  $i$  and  $j$ , where  $n_i$  = number of samples in class  
604  $i$  and  $n_j$  = number of samples in class  $j$  (see **Supplementary Figure 5a** for an example). We  
605 averaged the PRED values over all these matrices for each pair of classes and then computed  
606 the final PRED value by averaging over all class pairs.

607

## 608 **Other metrics for vector similarity in class-vector data**

609 PRED was compared to 5 other metrics of vector similarity: Pearson's correlation (PC),  
610 Cosine similarity (COS), Manhattan distance (MAN), Euclidean distance (EUC), and  
611 Chebyshev's distance (CHEB). If the dataset included more than two vectors, each of the  
612 metrics was calculated over all possible pairs of vectors and then averaged. PC was computed  
613 using the corr function in MATLAB; while analyzing experimental datasets, any rows with  
614 incomplete data were removed. COS was as 1 - cosine distance using the cosine option of  
615 the pdist function in MATLAB. The distance-based metrics MAN, EUC, and CHEB were  
616 calculated using the pdist function with the options cityblock, euclidean, and chebychev,  
617 respectively. Since the range of the distance-based metrics (MAN, EUC, and CHEB) was  
618 between 0 and  $\infty$ , we transformed these metrics using the negative exponential function  
619  $f(x) = e^{-x}$  which mapped the range to be between 1 and 0 such that a value close to 1  
620 indicated a small distance (high similarity) between the vectors.

621

## 622 **Other metrics for class separability in class-sample data**

623 PRED was compared to 5 other metrics of class separability: Euclidean template matching  
624 (ETM), Silhouette index (SIL), Davies-Bouldin index (DBI), Dunn's index (DUNN), and  
625 Calinski-Harabasz index (CH). ETM is based on a simple algorithm for calculating  
626 classification accuracy (Stopfer et al., 2003). Briefly, a template was created for each class by  
627 averaging the values within the class, excluding the test sample. Next, for each sample in the  
628 dataset, the Euclidean distances between the sample and all the templates were calculated. If  
629 the smallest distance belongs to the template of the actual class of the sample, the sample was  
630 correctly classified and scored as 1 (if templates of n classes, including the actual class of the  
631 sample, had the same smallest distance, the score was set to  $\frac{1}{n}$ ). Otherwise, the score was set  
632 to 0. The average of the scores from all the samples was reported as the final value of ETM.  
633 ETM ranges between 0 and 1, where 1 denotes the highest level of class separability (every  
634 sample is correctly classified). We used a custom function written in MATLAB for  
635 calculating the ETM values. The Silhouette index compares the pairwise intra-class and inter-  
636 class distances (Rousseeuw, 1987). It ranges between 1 and -1, where 1 indicates high  
637 separability. DBI is calculated as the ratio of within-class and between-class distances  
638 (Davies and Bouldin, 1979). It ranges from 0 to  $\infty$ , where 0 indicates high separability. CH  
639 measures the ratio of the average intra-class and inter-class variances (Caliński and Harabasz,  
640 1974). It ranges between 0 and  $\infty$ , where a higher value indicates higher separability. SIL,  
641 DBI, and CH were calculated using the evalclusters function in MATLAB, with the options  
642 Silhouette, DaviesBouldin, and CalinskiHarabasz, respectively. DUNN calculates the ratio  
643 of the minimum inter-cluster distance to the maximum intra-cluster distance (Dunn, 1974). It  
644 ranges between 0 and  $\infty$ , where a higher value indicates high separability. We calculated the  
645 DUNN value using the indexDN function written by Julian Ramos for MATLAB.

646

## 647 **Simulations with clusters of points**

648 To simulate a class-sample dataset, we first selected the class means uniformly distributed  
649 within an  $n$ -dimensional space  $[-1, 1]^n$ . The samples were then drawn from a uniform  
650 distribution around the class mean such that the Euclidean distance between the sample and  
651 the class mean was  $\leq r$ , where  $r$  denotes the cluster radius. Next, a random noise  $n$ -  
652 dimensional vector, drawn from  $[\mathcal{N}(0, \sigma)]^{1 \times n}$ , was added to each sample (see  
653 **Supplementary Figure 3f—h** for examples). Note that after the addition of noise, the  
654 samples no longer lay within  $[-1, 1]^n$  but, instead, within  $[-\infty, \infty]^n$ .

655

## 656 **Chance level**

657 The chance level for each metric was calculated using datasets with no inherent similarity or  
658 separability. For the class-vector metrics, we simulated a dataset of 10 vectors and either 2 or  
659 5 classes. Each value within the dataset was randomly drawn from a uniform distribution  
660 between -1 and 1, ensuring no structure within the classes or the vectors. The whole  
661 simulation was repeated 1000 times, and the vector similarity metrics were reported. For the  
662 class-sample metrics, we simulated a 2-dimensional clustered dataset with 10 samples and  
663 either 2 or 5 classes. The cluster radius was set to 0.05 for all the classes, and a big noise term  
664 randomly drawn from  $[\mathcal{N}(0, 50)]^{1 \times 2}$  was added to simulate inseparable clusters. The whole  
665 simulation was repeated 1000 times, and the class separability metrics were reported.

666

## 667 **Dynamic range and variability**

668 The dynamic range was defined as the range of noise levels in which a metric remains  
669 informative (i.e., does not saturate near the maximum or the minimum level). We simulated a  
670 dataset with increasing levels of noise (on a log scale). We measured the average value  
671 reported by the metric at the 5 lowest noise levels (as  $\mu(v_l)$ ) and at the 5 highest noise levels  
672 (as  $\mu(v_h)$ ) simulated. The absolute difference between these two values,  $|\mu(v_l) - \mu(v_h)|$ ,  
673 was called the vertical range of the metric. For a metric whose value decreased with  
674 increasing noise, the left boundary of the dynamic range was taken as the lowest noise level  
675 at which the average value of the metric was lower than the value at the lowest noise level by  
676 at least 1% of the vertical range, i.e.,  $DR_l = \min(x) : \mu(x) < \mu(v_l) - 0.01 \times |\mu(v_l) -$   
677  $\mu(v_h)|$ . The right boundary of the dynamic range was taken as the highest noise level at  
678 which the average metric value was greater than the value at the highest noise level tested  
679 plus 1% of the vertical range, i.e.,  $DR_h = \max(x) : \mu(x) > \mu(v_h) + 0.01 \times |\mu(v_l) - \mu(v_h)|$ .  
680 The dynamic range was calculated as  $|DR_h - DR_l|$ .

681 The variability of the metric was defined as the standard deviation of the metric at the mid-  
682 point of the dynamic range divided by its vertical range, i.e.,



683

$$\text{Variability} = \frac{\sigma\left(\frac{|DR_h + DR_l|}{2}\right)}{|\mu(v_l) - \mu(v_h)|}$$

684 where  $\sigma(x)$  represents the standard deviation in the metric values at the noise level  $x$ . For the  
685 class-vector metrics, we simulated a dataset with 10 vectors and 2 classes. The mean response  
686 of each class was set to 2 and 4, respectively. The value for a class was randomly drawn from  
687  $\mathcal{N}(\mu, \sigma)$ , where  $\mu$  is the class mean,  $\sigma = 10^v$  and  $v \in [-2, -1.9, -1.8, \dots, 3]$  to simulate  
688 increasing noise levels on a log scale, covering 5 orders of magnitude. Each simulation was  
689 repeated 1000 times, and the resultant similarity was measured using each metric. We  
690 repeated the entire experiment with increasing base means, i.e., we added an integer value to  
691 the mean response of the classes. For example, adding 1 to the class means changed them  
692 from [2 4] to [3 5]. We simulated 11 such datasets by adding each of the integers in the range  
693 [0 10].

694 For class-sample datasets, we simulated a dataset with 2 classes, each with 10 samples. The  
695 response was set as a 2-dimensional vector. The class means were drawn from the 2-D space  
696  $[-1 \ 1]^2$  with a cluster radius of 0.05. The noise was drawn randomly from  $\mathcal{N}(0, \sigma)$ , where  
697  $\sigma = 10^v$  and  $v \in [-3, -2.9, -2.8, \dots, 3]$  to simulate increasing noise levels (on a log scale)  
698 within the dataset. Each simulation was repeated 1000 times.

699 In the analysis where we added noise to the experimental data, we first calculated the mean  
700 response over all the different trials and odors ( $m$ ). The noise ( $v$ ) was then added to each  
701 value of the data matrix as a percentage of this mean response with the values drawn from  
702  $\mathcal{N}(0, \sigma)$ , where  $\sigma = 10^v \times m \times 0.01$ ,  $m$  is the mean response, and  $v \in$   
703  $[-1, -0.9, -0.8, \dots, 4]$  is the noise level on a log scale.

704

## 705 ***Drosophila* olfactory behavior**

706 We used a published dataset containing the behavioral preferences of 141 wild-type  
707 *Drosophila* for 3-octanol (OCT) versus 4-methylcyclohexanol (MCH) (Honegger et al.,  
708 2020). The behavior was quantified as a preference index obtained from a two-choice assay  
709 where the odors were presented, one on each port. A value above 0.5 indicated preference  
710 towards MCH while a value between 0 and 0.5 indicated preference towards OCT. The  
711 preferences were calculated for all the flies at two different time points, 24-hrs apart. We first  
712 calculated the individual-time (class-vector) PRED and Pearson's correlation (PC) values  
713 over the entire dataset (**Figure 2a**). To compare the stability of the two metrics for  
714 incomplete data, we randomly sampled 70 out of 141 individuals from the dataset. We  
715 calculated the PRED and PC value for this subset, repeating the random sampling 20 times.  
716 We then calculated the coefficient of variation of each metric over these 20 random  
717 samplings. To check the validity of our results, we repeated this entire process 50 times and  
718 compared the coefficient of variation obtained from the two metrics.

719

## 720 ***Drosophila* population responses**

721 To analyze the population level similarity in responses, we used a published dataset of  
722 calcium imaging responses of 37 glomeruli responding to 36 monomolecular odors (Badel et  
723 al., 2016). The glomeruli measured within the dataset were DM6, DM5, DM2, DM1, DM4,  
724 VM2, VM7d, VM7v, DA4L, DA2, DL1, DL5, D, DM3, DC2, VA6, DC3, DL4, DA3, DL3,  
725 DA1, VA1d, VA1v, VL2a, VL2p, VA5, VM4, VA7L, VA3, VA4, VA7m, VC2, VC1, VM3,  
726 VA2, VM1, and Dp1m. The odors used in the dataset were apple cider vinegar, mango  
727 mimic, broth, benzaldehyde, 2-methyl phenol, butanol, g-butyrolactone, methanoic acid,  
728 hexanoic acid, 1-octanol, acetophenone, vinegar mimic, 2,3-butanedione, pentanoic acid, 3-  
729 methylthio-1-propanol, 3-octanol, ethyl butyrate, 4-methylcyclohexanol, acetaldehyde, 2-  
730 pentanone, 2-oxopentanoic acid, hexyl acetate, isopentyl acetate, phenylethylamine,  
731 propionic acid, geosmin, ethyl acetate,  $\beta$ -citronellol, benzyl alcohol, linalool, 1-octen-3-ol,  
732 methyl salicylate, pentyl acetate, banana essence, 2-butanone, and 1-butanol. The dataset  
733 included the responses for 61 individuals (although not all individuals were measured for all  
734 odors) with around 4 trials each. For calculating the similarity within the individuals, we first  
735 averaged the responses over the trials. We then calculated the odor-individual (class-vector)  
736 PRED for each of the 37 different glomeruli separately (**Figure 3e**). Alternatively, we used  
737 the 37-length vectors as the values in the 61 (odor)  $\times$  36 (individual) matrix and calculated a  
738 single odor-individual (class-vector) PRED for these ‘population’ responses.

739

## 740 **Zebrafish olfactory behavior**

741 We extracted the published data of seven behavioral responses of 10 wild-type Zebrafish in  
742 response to 18 different odors over 4 different trials from the raw data files provided by the  
743 authors (Kermen et al., 2020). The odors for which the response of the zebrafish was tested  
744 were food extract (food), histidine (his), nucleotides (nucl), methionine (met), phenylalanine  
745 (phe), cysteine (cys), arginine (arg), bile acids (bile), prostaglandin 2 $\alpha$  (pgf2a), urea,  
746 ammonium (amo), putrescine (put), spermine (sperm), cadaverine (cad), chondroitin sulfate  
747 (cs), zebrafish blood (blood), zebrafish skin extract (skin), and artificial fish water (afw). The  
748 behaviors extracted were fish velocity, freezing behavior, vertical position in the arena,  
749 percentage of burst swimming, number of abrupt turns, number of horizontal swimming  
750 events, and number of vertical swimming events. We used custom scripts and MATLAB  
751 functions provided through personal correspondence by Dr. Florence Kermen to extract the  
752 data using the protocol described in the original paper (Kermen et al., 2020).

753 To characterize the individual-to-individual separability, we calculated the individual-time  
754 (class-sample) PRED value separately for each of the 18 odors. For each odor, the dataset  
755 included 10 classes (individuals) with 4 samples (trials) per class. The value of each sample  
756 was a 7-dimensional vector, representing the 7 behaviors (**Figure 7a**).

757

## 758 ***Drosophila* projection neuron responses with and without serotonin** 759 **blockage**

760 We obtained the published calcium imaging responses of 14 different projection neurons  
761 (PNs) from 18 different GCaMP6m wild-type flies and 7  $\alpha$ -methyl tryptophan (a-mw) fed  
762 flies to 12 different monomolecular odors (Honegger et al., 2020). The PNs in this dataset  
763 innervated DA1, DL3, DL1, DL5, DM3, DM6, DA2, DA4l, D, DM5, DM2, DM1, DM4, and  
764 DL4 glomeruli. The odors within the dataset were 3-octanol, 1-hexanol, ethyl-lactate,  
765 citronella, 2-heptanone, 1-pentanol, ethanol, geranyl-acetate, hexyl-acetate, 4-  
766 methylcyclohexanol, pentyl-acetate, 1-butanol. Each response was measured over 2 trials. We  
767 calculated individual-trial (class-sample) PRED separately for each PN-odor combination  
768 (**Figure 7c**).

769

## 770 **Locust and *Drosophila* electrophysiological recordings**

771 We used published recordings of the response of bLN1 mushroom body output neurons in 6  
772 different locusts responding to 6 different odors (Gupta and Stopfer, 2014). These  
773 electrophysiological responses were measured in awake locusts exposed to cyclohexanone,  
774 octanol, and hexanol in concentrations of 0.1% and 10% each. Each response consisted of 6-  
775 10 trials.

776 We also used the published responses of *Drosophila* PNs innervating 4 different glomeruli  
777 (VC4, DL2v, VM5v, VC3) to a set of 5 odors – benzaldehyde, 2-octanone, pentyl acetate,  
778 ethyl acetate, and ethyl butyrate (Shimizu and Stopfer, 2017) – although not all PNs were  
779 measured for all the odors. The response of each PN was measured in 2-6 individuals with  
780 approximately 6-10 trials per response.

781 For analyzing the odor-individual (class-vector) PRED with temporal responses, we extracted  
782 both the firing rate and the temporal response of the neurons for a period of 2-s after odor  
783 onset. The firing rate was calculated as the total number of spikes within the 2 second period  
784 from 2 to 4 seconds in the response minus the number of spikes in the 2 second period before  
785 odor onset, from 0 to 2 seconds in the response. The temporal response was similarly  
786 calculated in the 2 second period after odor onset divided into 10-bins of 200 ms each minus  
787 one-tenth the total number of spikes in the background response from 0 to 2 seconds. For  
788 calculating PRED and PC, we first averaged the responses over all the trials for each cell in  
789 the dataset (1 cell in the locust dataset and 4 cells in the *Drosophila* dataset). We then  
790 calculated the odor-individual (class-vector) PRED using both the firing rate (magnitude) and  
791 the temporal responses. PRED values were averaged over all pairs of odors for every pair of  
792 individuals.

793 In the experiments where we added noisy bins to the experimental datasets, we used the  
794 initial 10-bin vector of responses as the base dataset. For adding one noisy bin to the base  
795 dataset, we used the number of spikes obtained from 4 to 4.2 seconds minus the background

796 response as the eleventh bin. Similarly, any extra noise bin extended the response period by  
797 200 ms to a maximum of 4 seconds when 10 extra noise bins were added.

798 In the experiment, where we investigated the applicability of PRED to class-sample datasets,  
799 we used both the locust and the fly databases to calculate odor-trial (class-sample) PRED and  
800 compared it to the odor separability obtained from the other metrics. For each individual and  
801 cell in the dataset, we used the 2-bin (each bin of length 1 second) response vector to  
802 calculate the separability.

803

## 804 **Temporal response simulations**

805 To simulate the temporal responses, we created a dataset with 2 classes and 10 vectors, where  
806 each response was a 10-bin vector. The base mean of each response bin within a class was  
807 randomly drawn from a uniform distribution in the range [1 3]. A random noise drawn from  
808  $\mathcal{N}(0,1)$  was added to each bin. A particular number of extra bins were appended to the  
809 vectors, with each new bin containing a value with a base mean of 0 and a noise drawn from  
810  $\mathcal{N}(0,1)$ . We compared the PRED and PC values with the number of extra bins ranging from  
811 0 to 10. The entire simulation was repeated 100 times. To further emphasize the difference  
812 between the behaviors of PRED and PC, we repeated this entire simulation by generating  
813 extra bins that were exactly 0 (without any noise).

814

## 815 **Simulations with increasing numbers of classes or samples**

816 We generated 2-dimensional clustered data with cluster means drawn from  $[-1\ 1]^{1 \times 2}$  and  
817 cluster radius of 0.05. A small amount of noise drawn from  $\mathcal{N}(0, 0.4)^{1 \times 2}$  was added to each  
818 response in the dataset. For the simulations with increasing numbers of classes, we simulated  
819 two different datasets – one with 2 samples and the other with 10 samples. The number of  
820 classes ranged from 2 to 10. For the simulations with increasing numbers of samples, we used  
821 2 classes. The number of samples was taken from [2, 4, ..., 20]. Each simulation was  
822 repeated 100 times.

823 For comparing ‘fast PRED’ with ‘exhaustive PRED’, we used the same dataset of 2 classes as  
824 described above but varied the number of samples from [2, 3, ..., 25]. Each simulation was  
825 repeated 1000 times. The average value of PRED over all simulations was  $\sim 0.5$ . For each  
826 simulation and number of samples, we calculated the absolute difference between ‘fast  
827 PRED’ and ‘exhaustive PRED’ values. Finally, we reported the average difference over  
828 simulations divided by the average ‘exhaustive PRED’ value for the specified number of  
829 samples.

830

## 831 ***Drosophila* connectome data**

832 We obtained the connectivity vectors of identified local horn neurons (LHNs) from Schlegel  
833 et al. (Schlegel et al., 2021) for 87 identified neurons in the FAFB and the Hemibrain  
834 databases. The dataset we used included 47 neurons from FAFB and 85 neurons from  
835 Hemibrain along with their connectivity to 57 unique antennal lobe glomeruli (D, DA1, DA2,  
836 DA3, DA4l, DA4m, DC1, DC2, DC3, DC4, DL1, DL2d, DL2v, DL3, DL4, DL5, DM1,  
837 DM2, DM3, DM4, DM5, DM6, DP1l, DP1m, V, VA1d, VA1v, VA2, VA3, VA4, VA5,  
838 VA6, VA7l, VA7m, VC1, VC2, VC3, VC4, VC5, VL1, VL2a, VL2p, VM1, VM2, VM3,  
839 VM4, VM5d, VM5v, VM6, VM7d, VM7v, VP1d, VP1l, VP1m, VP2, VP3, VP5). The LHNs  
840 were grouped into 49 ‘connectivity types,’ which were further grouped into 36 ‘cell types’,  
841 then 13 ‘tracts’, and finally 4 ‘regions’, based on their morphologies within the lateral horn  
842 (Frechter et al., 2019; Schlegel et al., 2021).

843 The full dataset consisted of unique connectivity types as classes and the two databases as  
844 vectors. The connectivity vector of each neuron within a connectivity type was averaged.  
845 Each cell within this matrix was a 57-length vector of averaged and normalized connectivity  
846 weights of the corresponding LHN to each glomerulus. We first calculated the connectivity  
847 type-database (class-vector) PRED value over this matrix to characterize the similarity of  
848 connections across databases. Next, we grouped this matrix based on each of the different  
849 hierarchy levels. We averaged the connectivity vectors over all connectivity types belonging  
850 to a group within a particular hierarchy to get a matrix with groups as columns and the  
851 databases as rows (**Figure 8a**). We then calculated the group-database (class-vector) PRED  
852 values for each hierarchy level based on cell type, tract, or region.

853 In the experiment where we characterized the separability of neurons across groups based on  
854 their connectivity to antennal lobe glomeruli, we constructed 4 different matrices with  
855 individual neurons (not averaged over connectivity types) as samples and the relevant group  
856 types as classes for the two databases separately (**Figure 8b**). We then calculated the group-  
857 neuron (class-sample) PRED for each matrix to characterize the separability of neural  
858 connectivity vectors across groups for each hierarchy level.

859

## 860 **Statistics**

861 To compare a set of PRED values with the baseline (0) or a specific mean, we used a one-  
862 sample double-sided t-test. To compare the chance level of the metrics across classes, we  
863 used two-sample double-sided unpaired t-tests. For comparing the coefficient of variation  
864 obtained for PRED with those for PC, we used a two-sample double-sided paired t-test.

865



## 866 **Code availability**

867 All the simulations and analyses were done using custom scripts coded in MATLAB (version  
868 r2020a). A modified version of the *gramm* plotting package (Morel, 2018) was used for all  
869 the figure plots. The source code for the simulations and analysis can be found at  
870 [https://github.com/neuralsystems/PRED\\_analysis](https://github.com/neuralsystems/PRED_analysis). The standalone versions of PRED function  
871 written in Python and MATLAB can be found at <https://github.com/neuralsystems/PRED>  
872 (the MATLAB version is also available on the MATLAB File Exchange).

873

## 874 **Competing interests**

875 The authors declare no competing interests.

876

## 877 **Acknowledgments**

878 We thank Kazumichi Shimizu and Mark Stopfer for sharing fly electrophysiology data and  
879 Hokto Kazama for sharing fly calcium imaging data. We thank Florence Kermen and  
880 Alexander Bates for helping us extract the zebrafish behavioral data and the connectomics  
881 data, respectively. We thank Arjit Kant Gupta for providing an initial version of the Python  
882 implementation for PRED. This work was supported by the DBT/Wellcome Trust India  
883 Alliance Fellowship [grant number IA/I/15/2/502091] awarded to N.G.; Cognitive Science  
884 Research Initiative of the Department of Science & Technology [DST/CSRI/2018/102] to  
885 N.G.; SERB Core Research Grant [CRG/2020/004719] to N.G.; a BBSRC grant  
886 [BB/S016031/1] to A.L.; and a Starting Grant from the European Research Council [639489]  
887 to A.L.

888

## 889 **Author Contributions (using CRediT format)**

890 A.M.M. Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology,  
891 Software, Visualization, Writing – original draft, Writing – review & editing

892 A.C.L. Conceptualization, Funding acquisition, Methodology, Writing–review and editing

893 N.G. Conceptualization, Formal analysis, Funding acquisition, Investigation, Project  
894 administration, Supervision, Visualization, Methodology, Writing–original draft, Writing–  
895 review and editing

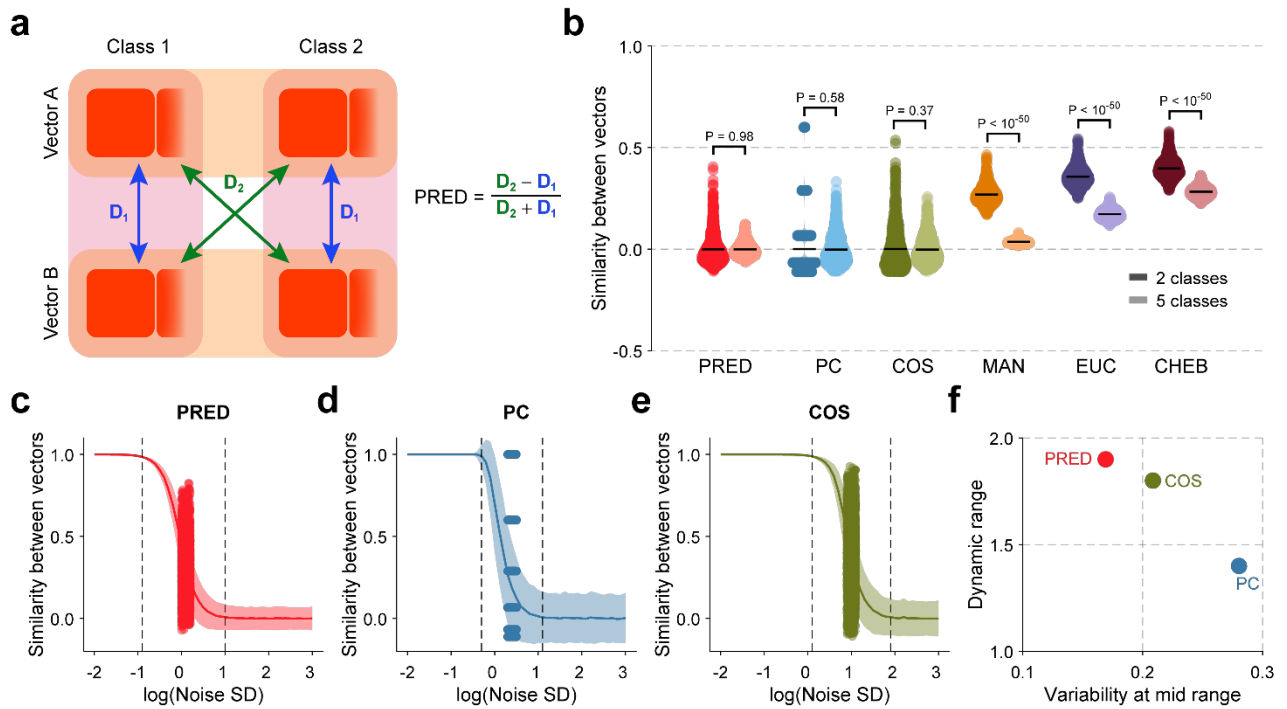
## 896 References

- 897 Arbelaitz O, Gurrutxaga I, Muguerza J, Pérez JM, Perona I. 2013. An extensive comparative  
898 study of cluster validity indices. *Pattern Recognition* **46**:243–256.  
899 doi:10.1016/j.patcog.2012.07.021
- 900 Badel L, Ohta K, Tsuchimoto Y, Kazama H. 2016. Decoding of Context-Dependent  
901 Olfactory Behavior in *Drosophila*. *Neuron* **91**:155–167.  
902 doi:10.1016/j.neuron.2016.05.022
- 903 Brun M, Sima C, Hua J, Lowey J, Carroll B, Suh E, Dougherty ER. 2007. Model-based  
904 evaluation of clustering validation measures. *Pattern Recognition* **40**:807–824.  
905 doi:10.1016/j.patcog.2006.06.026
- 906 Buchanan SM, Kain JS, Bivort BL de. 2015. Neuronal control of locomotor handedness in  
907 *Drosophila*. *PNAS* **112**:6700–6705. doi:10.1073/pnas.1500804112
- 908 Caliński T, Harabasz J. 1974. A dendrite method for cluster analysis. *Communications in*  
909 *Statistics* **3**:1–27. doi:10.1080/03610927408827101
- 910 Davies DL, Bouldin DW. 1979. A Cluster Separation Measure. *IEEE Transactions on*  
911 *Pattern Analysis and Machine Intelligence* **PAMI-1**:224–227.  
912 doi:10.1109/TPAMI.1979.4766909
- 913 Dunn JC. 1974. Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of*  
914 *Cybernetics* **4**:95–104. doi:10.1080/01969727408546059
- 915 Frechter S, Bates AS, Tootoonian S, Dolan M-J, Manton JD, Jamasb AR, Kohl J, Bock D,  
916 Jefferis GS. 2019. Functional and anatomical specificity in a higher olfactory centre.  
917 *eLife* **8**:e44590. doi:10.7554/eLife.44590
- 918 Guerra L, Robles V, Bielza C, Larrañaga P. 2012. A comparison of clustering quality indices  
919 using outliers and noise. *IDA* **16**:703–715. doi:10.3233/IDA-2012-0545
- 920 Gupta N, Stopfer M. 2014. A Temporal Channel for Information in Sparse Sensory Coding.  
921 *Current Biology* **24**:2247–2256. doi:10.1016/j.cub.2014.08.021
- 922 Gurrutxaga I, Muguerza J, Arbelaitz O, Pérez JM, Martín JI. 2011. Towards a standard  
923 methodology to evaluate internal cluster validity indices. *Pattern Recognition Letters*  
924 **32**:505–515. doi:10.1016/j.patrec.2010.11.006
- 925 Honegger KS, Smith MA-Y, Churgin MA, Turner GC, Bivort BL de. 2020. Idiosyncratic  
926 neural coding and neuromodulation of olfactory individuality in *Drosophila*. *PNAS*  
927 **117**:23292–23297. doi:10.1073/pnas.1901623116
- 928 Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional  
929 architecture in the cat's visual cortex. *The Journal of Physiology* **160**:106–154.  
930 doi:10.1113/jphysiol.1962.sp006837
- 931 Huerta R, Nowotny T, García-Sánchez M, Abarbanel HDI, Rabinovich MI. 2004. Learning  
932 Classification in the Olfactory System of Insects. *Neural Computation* **16**:1601–1640.  
933 doi:10.1162/089976604774201613
- 934 Kain JS, Zhang S, Akhund-Zade J, Samuel ADT, Klein M, de Bivort BL. 2015. Variability in  
935 thermal and phototactic preferences in *Drosophila* may reflect an adaptive bet-  
936 hedging strategy. *Evolution* **69**:3171–3185. doi:10.1111/evo.12813
- 937 Karagiannis A, Gallopin T, Dávid C, Battaglia D, Geoffroy H, Rossier J, Hillman EMC,  
938 Staiger JF, Cauli B. 2009. Classification of NPY-Expressing Neocortical  
939 Interneurons. *J Neurosci* **29**:3642–3659. doi:10.1523/JNEUROSCI.0058-09.2009
- 940 Kermen F, Darnet L, Wiest C, Palumbo F, Bechert J, Uslu O, Yaksi E. 2020. Stimulus-  
941 specific behavioral responses of zebrafish to a large range of odors exhibit individual  
942 variability. *BMC Biology* **18**:66. doi:10.1186/s12915-020-00801-8

- 943 Linneweber GA, Andriatsilavo M, Dutta SB, Bengochea M, Hellbruegge L, Liu G, Ejsmont  
944 RK, Straw AD, Wernet M, Hiesinger PR, Hassan BA. 2020. A neurodevelopmental  
945 origin of behavioral individuality in the *Drosophila* visual system. *Science* **367**:1112–  
946 1119. doi:10.1126/science.aaw7182
- 947 Lovmar L, Ahlford A, Jonsson M, Syvänen A-C. 2005. Silhouette scores for assessment of  
948 SNP genotype clusters. *BMC Genomics* **6**:35. doi:10.1186/1471-2164-6-35
- 949 McLean M, Stuart-Smith RD, Villéger S, Auber A, Edgar GJ, MacNeil MA, Loiseau N,  
950 Leprieur F, Mouillot D. 2021. Trait similarity in reef fish faunas across the world's  
951 oceans. *PNAS* **118**. doi:10.1073/pnas.2012318118
- 952 Mittal AM, Gupta D, Singh A, Lin AC, Gupta N. 2020. Multiple network properties  
953 overcome random connectivity to enable stereotypic sensory responses. *Nat Commun*  
954 **11**:1–15. doi:10.1038/s41467-020-14836-6
- 955 Morel P. 2018. Gramm: grammar of graphics plotting in Matlab. *Journal of Open Source*  
956 *Software* **3**:568. doi:10.21105/joss.00568
- 957 Niemelä M, Äyrämö S, Kärkkäinen T. 2018. Comparison of cluster validation indices with  
958 missing data. Presented at the European Symposium on Artificial Neural Networks,  
959 Computational Intelligence and Machine Learning. ESANN.
- 960 Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de  
961 Rijn M, Waltham M, Pergamenschikov A, Lee JCF, Lashkari D, Shalon D, Myers  
962 TG, Weinstein JN, Botstein D, Brown PO. 2000. Systematic variation in gene  
963 expression patterns in human cancer cell lines. *Nat Genet* **24**:227–235.  
964 doi:10.1038/73432
- 965 Rossum MCW van. 2001. A Novel Spike Distance. *Neural Computation* **13**:751–763.  
966 doi:10.1162/089976601300014321
- 967 Rousseeuw PJ. 1987. Silhouettes: A graphical aid to the interpretation and validation of  
968 cluster analysis. *Journal of Computational and Applied Mathematics* **20**:53–65.  
969 doi:10.1016/0377-0427(87)90125-7
- 970 Sasai S, Koike T, Sugawara SK, Hamano YH, Sumiya M, Okazaki S, Takahashi HK, Taga  
971 G, Sadato N. 2021. Frequency-specific task modulation of human brain functional  
972 networks: A fast fMRI study. *NeuroImage* **224**:117375.  
973 doi:10.1016/j.neuroimage.2020.117375
- 974 Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura Shin-ya, Hayworth KJ, Huang GB,  
975 Shinomiya K, Maitlin-Shepard J, Berg S, Clements J, Hubbard PM, Katz WT,  
976 Umayam L, Zhao T, Ackerman D, Blakely T, Bogovic J, Dolafi T, Kainmueller D,  
977 Kawase T, Khairy KA, Leavitt L, Li PH, Lindsey L, Neubarth N, Olbris DJ, Otsuna  
978 H, Trautman ET, Ito M, Bates AS, Goldammer J, Wolff T, Svirskas R, Schlegel P,  
979 Neace E, Knecht CJ, Alvarado CX, Bailey DA, Ballinger S, Borycz JA, Canino BS,  
980 Cheatham N, Cook M, Dreher M, Duclos O, Eubanks B, Fairbanks K, Finley S,  
981 Forknall N, Francis A, Hopkins GP, Joyce EM, Kim S, Kirk NA, Kovalyak J, Lauchie  
982 SA, Lohff A, Maldonado C, Manley EA, McLin S, Mooney C, Ndama M, Ogundeyi  
983 O, Okeoma N, Ordish C, Padilla N, Patrick CM, Paterson T, Phillips EE, Phillips EM,  
984 Rampally N, Ribeiro C, Robertson MK, Rymer JT, Ryan SM, Sammons M, Scott  
985 AK, Scott AL, Shinomiya A, Smith C, Smith K, Smith NL, Sobeski MA, Suleiman A,  
986 Swift J, Takemura Satoko, Talebi I, Tarnogorska D, Tenshaw E, Tokhi T, Walsh JJ,  
987 Yang T, Horne JA, Li F, Parekh R, Rivlin PK, Jayaraman V, Costa M, Jefferis GS, Ito  
988 K, Saalfeld S, George R, Meinertzhagen IA, Rubin GM, Hess HF, Jain V, Plaza SM.  
989 2020. A connectome and analysis of the adult *Drosophila* central brain. *eLife*  
990 **9**:e57443. doi:10.7554/eLife.57443

- 991 Schlegel P, Bates AS, Stürner T, Jagannathan SR, Drummond N, Hsu J, Serratos Capdevila  
992 L, Javier A, Marin EC, Barth-Maron A, Tamimi IF, Li F, Rubin GM, Plaza SM, Costa  
993 M, Jefferis GS. 2021. Information flow, cell types and stereotypy in a full olfactory  
994 connectome. *eLife* **10**:e66018. doi:10.7554/eLife.66018
- 995 Shimizu K, Stopfer M. 2017. A Population of Projection Neurons that Inhibits the Lateral  
996 Horn but Excites the Antennal Lobe through Chemical Synapses in *Drosophila*.  
997 *Frontiers in Neural Circuits* **11**:30. doi:10.3389/fncir.2017.00030
- 998 Stopfer M, Jayaraman V, Laurent G. 2003. Intensity versus Identity Coding in an Olfactory  
999 System. *Neuron* **39**:991–1004. doi:10.1016/j.neuron.2003.08.011
- 1000 Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. 2019. High-dimensional  
1001 geometry of population responses in visual cortex. *Nature* **571**:361–365.  
1002 doi:10.1038/s41586-019-1346-5
- 1003 VanderWerf F, Brassinga P, Reits D, Aramideh M, Ongerboer de Visser B. 2003. Eyelid  
1004 Movements: Behavioral Studies of Blinking in Humans Under Different Stimulus  
1005 Conditions. *Journal of Neurophysiology* **89**:2784–2796. doi:10.1152/jn.00557.2002
- 1006 Victor JD, Purpura KP. 1997. Metric-space analysis of spike trains: theory, algorithms and  
1007 application. *Network: Computation in Neural Systems* **8**:127–164. doi:10.1088/0954-  
1008 898X\_8\_2\_003
- 1009 Victor JD, Purpura KP. 1996. Nature and precision of temporal coding in visual cortex: a  
1010 metric-space analysis. *Journal of Neurophysiology* **76**:1310–1326.  
1011 doi:10.1152/jn.1996.76.2.1310
- 1012 Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J,  
1013 Fisher CB, Sharifi N, Calle-Schuler SA, Kmecova L, Ali IJ, Karsh B, Trautman ET,  
1014 Bogovic JA, Hanslovsky P, Jefferis GSXE, Kazhdan M, Khairy K, Saalfeld S, Fetter  
1015 RD, Bock DD. 2018. A Complete Electron Microscopy Volume of the Brain of Adult  
1016 *Drosophila melanogaster*. *Cell* **174**:730-743.e22. doi:10.1016/j.cell.2018.06.019  
1017
- 1018

1019 **Figure 1**



1020

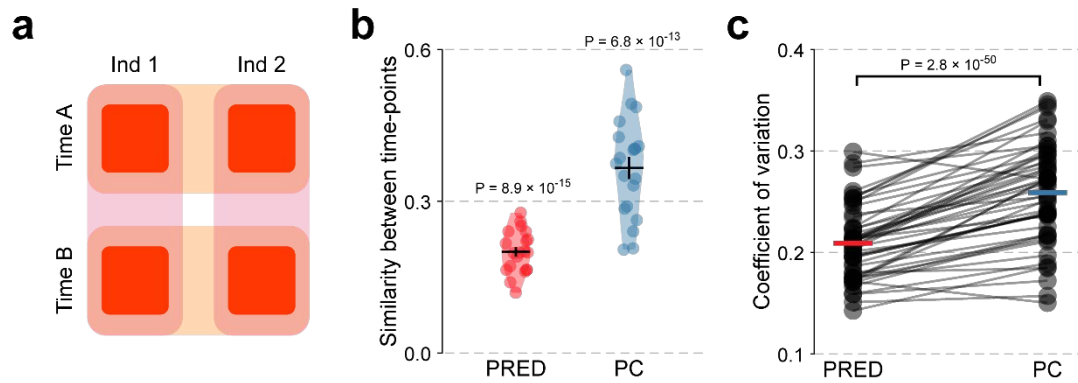
1021 **Figure 1: PRED is a robust metric for the assessment of similarity across vectors**

1022 **a** Schematic representation of Pairwise Relative Distance's (PRED) calculation for a class-  
 1023 vector dataset. **b** Violin plots showing the chance level of each metric with simulated datasets  
 1024 containing 2 (darker colors) or 5 (lighter colors) classes. Each point within a violin represents  
 1025 the metric's value for a different random seed (n = 1000 simulations for each number of  
 1026 classes). Note the change in the chance level of MAN, EUC, and CHEB metrics with the  
 1027 number of classes. PRED: Pairwise relative distance, PC: Pearson's correlation, COS: Cosine  
 1028 similarity, MAN: Manhattan distance, EUC: Euclidean distance, CHEB: Chebyshev's  
 1029 distance. Black horizontal line represents the mean. Error bars represent s.e.m. **c—e** Change  
 1030 in the value of PRED (c), PC (d), and COS (e) with increasing noise level (shown on a log  
 1031 scale) in a simulated dataset with 2 classes and 10 individuals. The dark line shows the mean  
 1032 value over all simulations at the specified noise level (n = 1000 simulations per noise level).  
 1033 The shaded area represents 1 standard deviation around the mean. The two dashed vertical  
 1034 lines represent the boundaries of the dynamic range. Each point represents a different random  
 1035 simulation at the noise level corresponding to the mid-point of the dynamic range. **f** The  
 1036 dynamic range and the variability at the mid-point of the dynamic range are shown for each  
 1037 metric. PRED showed the highest dynamic range and the lowest variability.

1038



1039 **Figure 2**



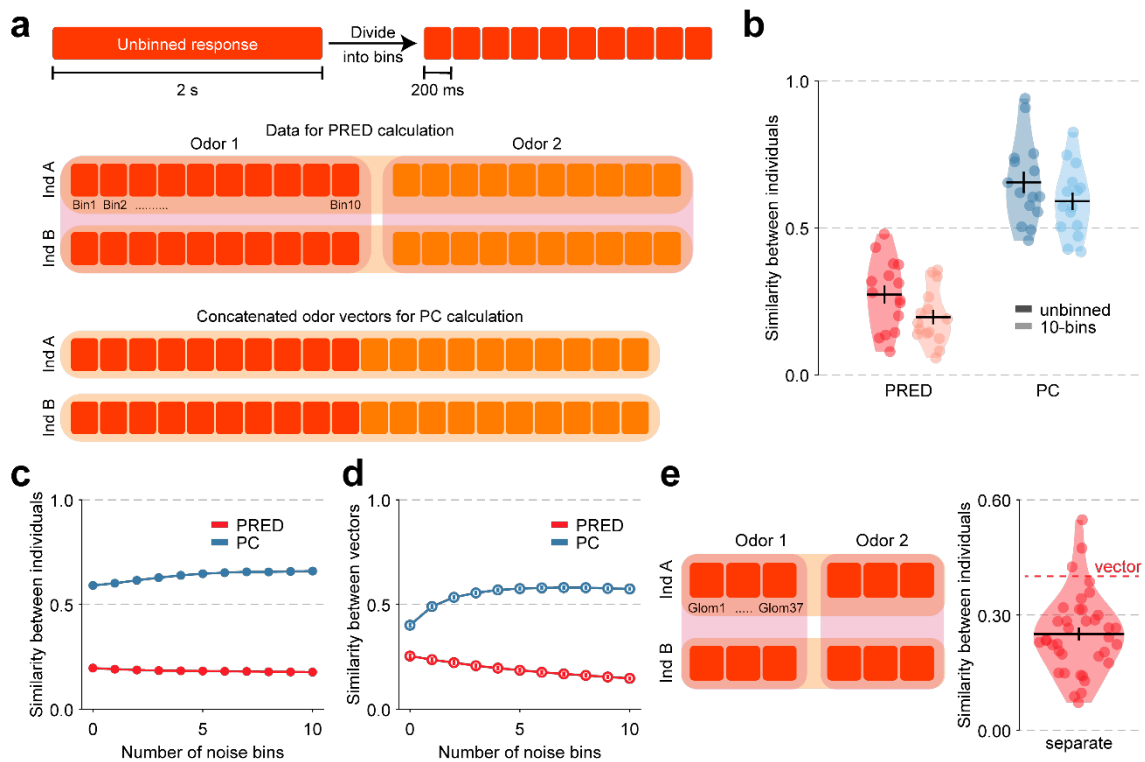
1040

1041 **Figure 2: PRED is a suitable metric for measuring behavioral similarity**

1042 **a** Illustration of an individual-time dataset where each value represents the preference index  
1043 of an individual animal at the specified time. **b** Across-time similarity in the MCH-OCT  
1044 preference index of *Drosophila* measured with 70 individuals and 2 time-points. The 70  
1045 individuals were randomly sampled from a dataset with 141 individuals. The coefficient of  
1046 variation (COV) is also displayed. Each point within a violin represents the mean similarity  
1047 for a new randomly sampled dataset ( $n = 20$  samplings). Black horizontal line represents the  
1048 mean. **c** Coefficient of variations of 100 different repetitions of the analysis performed in (**b**).  
1049 Horizontal lines represent the mean COV over all repetitions ( $n = 50$  repetitions). Lines  
1050 connect the PRED and PC values from the same repetition.

1051

1052 **Figure 3**

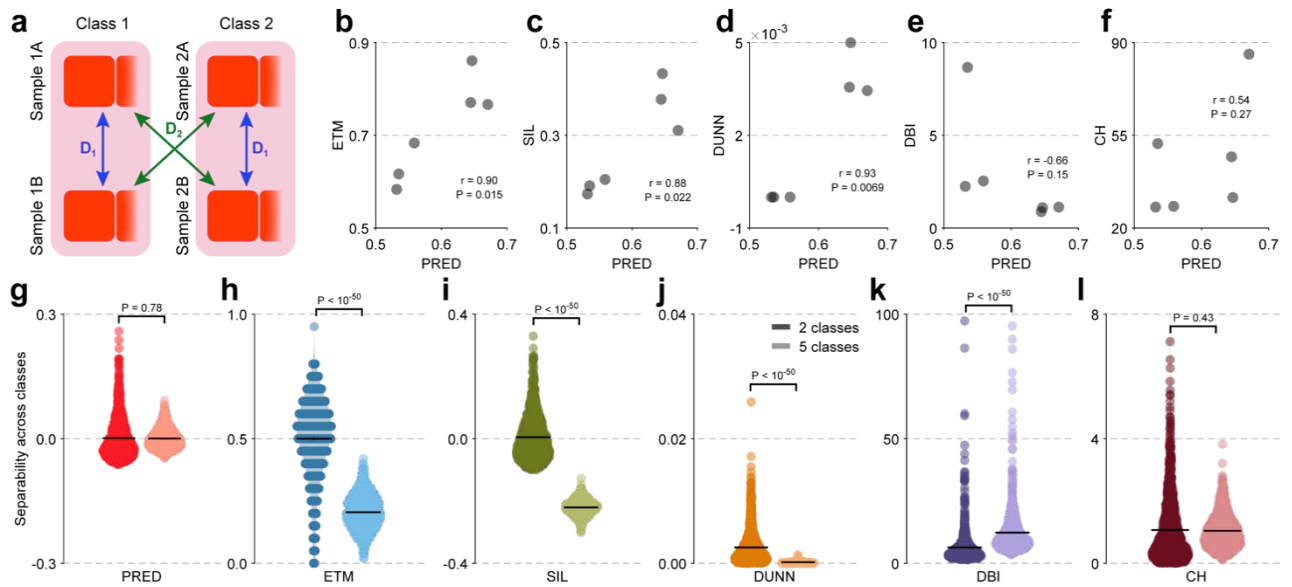


1053

1054 **Figure 3: PRED natively supports multi-dimensional data**

1055 **a** Illustrations showing the unbidden and the 10-bin temporal vectors used for calculating the  
 1056 response similarity between individuals. For calculating PRED, the Euclidean distance  
 1057 between the 10-bin vectors across individuals is calculated. However, for calculating PC, the  
 1058 responses for both odors are first concatenated into a single 20-bin vector and then correlated  
 1059 across individuals. **b** Across-individual similarity when the neural response is quantified as a  
 1060 single unbidden number (darker colors) or as a 10-bin temporal vector (lighter colors). The  
 1061 data is taken from locust bLN1 neural responses (Gupta et al. 2014). Each point within the  
 1062 violin represents the similarity for a pair of individuals ( $n = 15$ ). Black horizontal lines  
 1063 represent the mean, and error bars represent s.e.m. in all panels. **c** Across-individual  
 1064 similarity as a function of the number of extra bins (containing mostly noise) added to the  
 1065 original 10-bin vector for the same dataset as in (b). Note that the similarity value reported by  
 1066 PC increases with the increasing number of bins. **d** Across-individual similarity as a function  
 1067 of the number of extra bins (containing noise) added to a 10-bin vector for simulated data  
 1068 with 2 odors and 10 individuals. The value in each extra bin is taken from a normal  
 1069 distribution with 0 mean and 1 s.d. Open circles denote the mean over 100 different random  
 1070 simulations. The similarity gradually reduces with the increasing number of noisy bins for  
 1071 PRED but increases for PC. **e** Illustration of the odor-individual dataset used for comparing  
 1072 the population response across individuals. Each bin represents the response of a glomerulus  
 1073 (Glom) in an individual for the odor tested. Violin plot shows the across-individual similarity  
 1074 measured by odor-individual (class-vector) PRED in a database with a population of 37  
 1075 neurons, either considered separately (violin plot, where each point represents the PRED  
 1076 value for a neuron,  $n = 37$ ) or considered together as a population vector (red dashed line).

1077 **Figure 4**



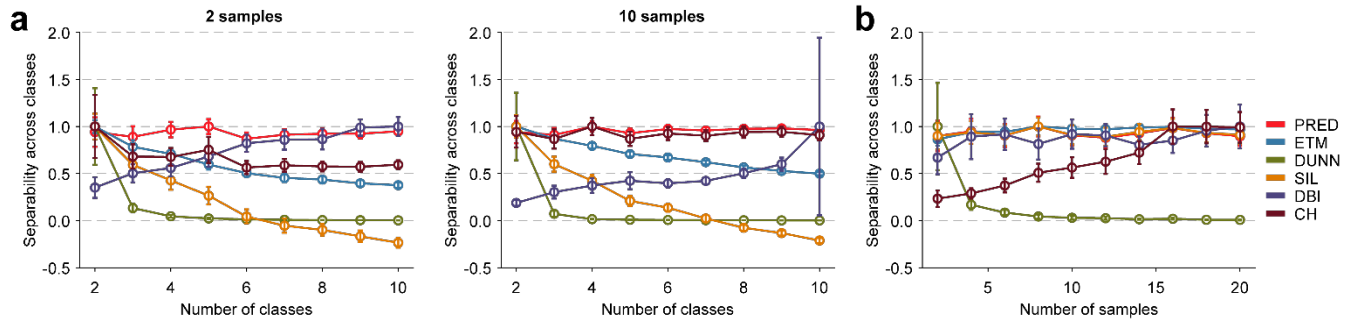
1078

1079 **Figure 4: PRED is suitable for assessing class separability in class-sample datasets**

1080 **a** Schematic representation of Pairwise Relative Distance (PRED) calculation for a class-  
 1081 sample dataset. **b—f** Odor separability measured using PRED compared to that measured  
 1082 using other commonly used metrics. Each point corresponds to one individual in the dataset  
 1083 taken from locust bLN1 neural responses (Gupta and Stopfer, 2014) (n = 6 individuals). Note  
 1084 that PRED values were positively correlated with the values obtained from other metrics  
 1085 (DBI expectedly showed a negative correlation as DBI's polarity is inverted). ETM:  
 1086 Euclidean template matching, SIL: Silhouette index, DUNN: Dunn's index, DBI: Davies-  
 1087 Bouldin index, CH: Calinski-Harabasz index. **g—l** Violin plots showing the chance level of  
 1088 each metric with simulated datasets containing 2 (darker colors) or 5 (lighter colors)  
 1089 classes. Each point within a violin represents the metric's value for a different random seed (n = 1000  
 1090 simulations for each number of classes). Note the change in the chance level of all metrics  
 1091 except PRED and CH with the number of classes. Black horizontal line represents the mean.  
 1092 Error bars represent s.e.m.

1093

1094 **Figure 5**



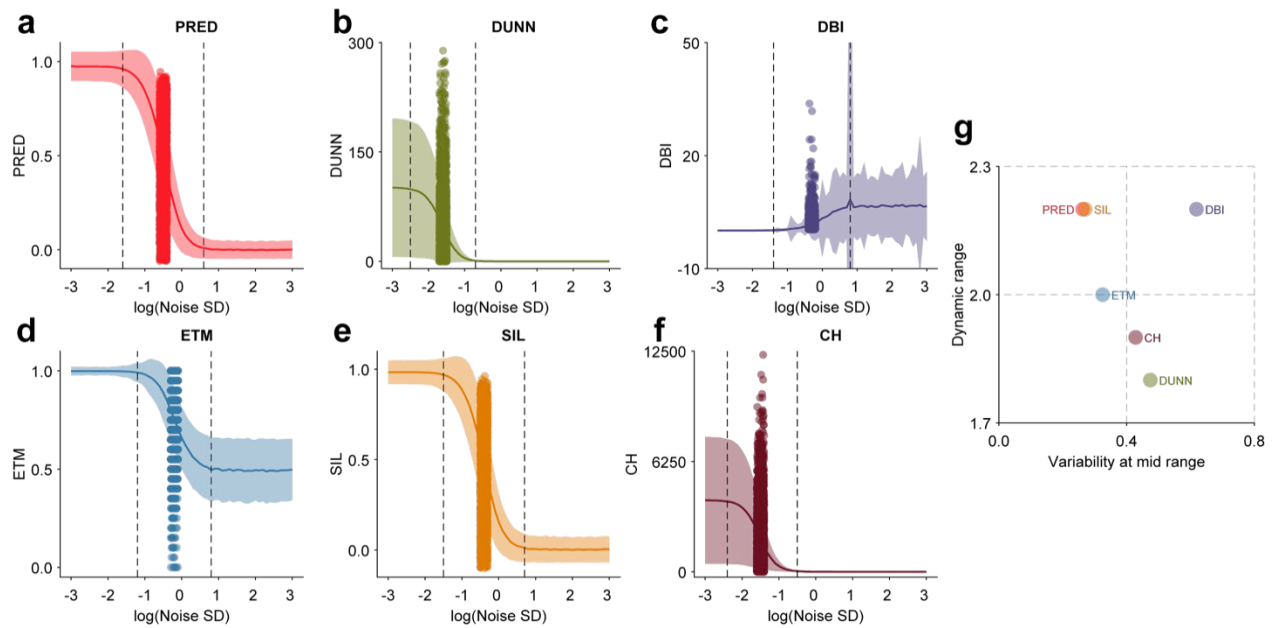
1095

1096 **Figure 5: Unlike PRED, most other metrics vary with an increasing number of classes or**  
1097 **samples**

1098 **a** Class separability as a function of the number of classes using simulated data with 2  
1099 samples (left) or 10 samples (right). Each metric was normalized by its maximum value  
1100 observed among the mean values for different numbers of classes. Note that all metrics  
1101 except PRED and SIL show change with the increasing number of classes. Open circles  
1102 denote the mean value over 100 different random simulations for the specified numbers of  
1103 classes, and error bars denote s.e.m. **b** Similar plot as in (a) but with 2 classes and an  
1104 increasing number of samples (n = 100 simulations for each number of samples). Note the  
1105 change in the value of ETM, DUNN, and CH with an increase in the number of samples.  
1106 Also, in all plots, DBI values show an opposite trend as compared with the other metrics  
1107 because DBI is higher for less separable classes.

1108

1109 **Figure 6**



1110

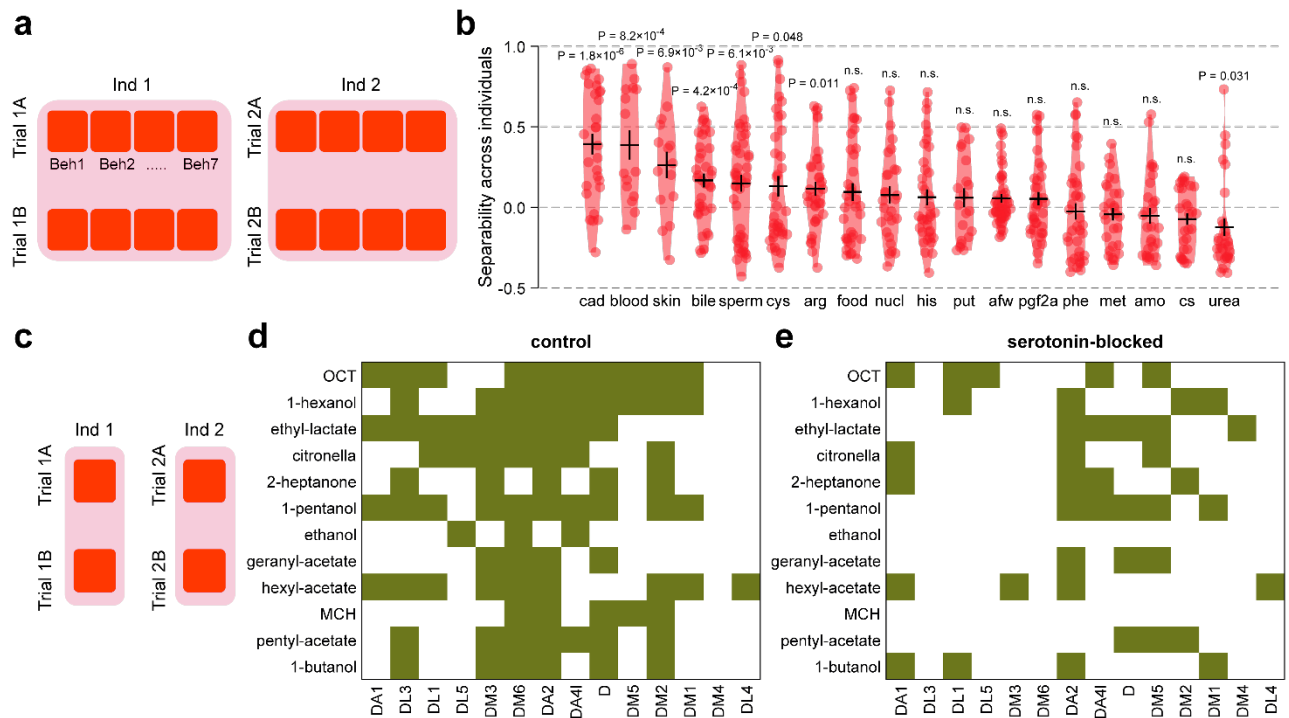
1111 **Figure 6: Comparison of dynamic range and variability of class-sample metrics**

1112 **a—f** Change in the value of PRED (a), DUNN (b), DBI (c), ETM (d), SIL (e), and CH (f)  
 1113 with increasing level of noise (shown on a log scale) in a simulated dataset with 3 classes and  
 1114 10 samples. The solid trace shows the mean values over all simulations for each noise level  
 1115 ( $n = 1000$  simulations per noise level). The shaded area represents 1 s.d. around the mean.  
 1116 The dashed vertical lines represent the boundaries of the dynamic range. Each point  
 1117 represents a different random simulation at the noise level corresponding to the mid-point of  
 1118 the dynamic range. **g** The dynamic range and the variability at the mid-point of the dynamic  
 1119 range are shown for each metric. PRED showed a reasonably large dynamic range and low  
 1120 variability.

1121



1122 **Figure 7**



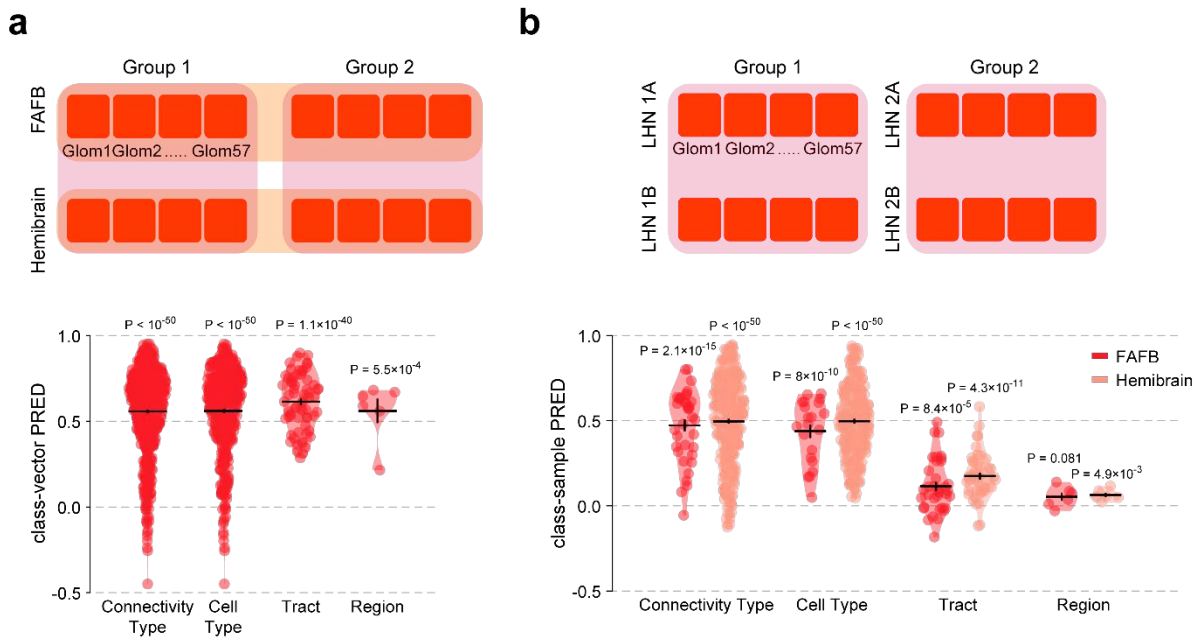
1123

1124 **Figure 7: Using PRED to measure individuality of neural responses**

1125 **a** Illustrations of an individual-trial dataset where values in each column represent the  
 1126 repeated behavioral responses of an individual (to a particular odor). Each behavioral  
 1127 response is a 7-length vector, with each bin in this vector representing a specific physical  
 1128 behavior (Beh). **b** Individual-trial (class-sample) PRED for zebrafish behavioral data  
 1129 calculated separately for each odor. The odors are sorted from left to right in decreasing order  
 1130 of PRED value. Each point in the violin represents an individual pair (cad:  $n = 28$ , blood:  $n =$   
 1131  $15$ , skin:  $n = 15$ , bile:  $n = 36$ , sperm:  $n = 45$ , cys:  $n = 36$ , arg:  $n = 36$ , food:  $n = 36$ , nucl:  $n =$   
 1132  $28$ , his:  $n = 36$ , put:  $n = 21$ , afw:  $n = 45$ , pgf2a:  $n = 36$ , phe:  $n = 36$ , met:  $n = 28$ , amo:  $n = 28$ ,  
 1133 cs:  $n = 28$ , urea:  $n = 28$ ). Black horizontal line represents the mean. Error bars represent  
 1134 s.e.m. n.s. means not significant. **c** Illustration of an individual-trial dataset where values in  
 1135 each column represent the repeated responses of an individual (in a particular glomerulus and  
 1136 to a particular odor). **d, e** Individual-trial (class-sample) PRED for different PN-odor  
 1137 responses in control (**d**) and serotonin-blocked (**e**) *Drosophila*. Green color indicates PRED  
 1138 values significantly greater than 0, indicating good separability across individuals. Note the  
 1139 fewer number of green values after serotonin-blockage. Significance was measured using  
 1140 one-sample t-test.

1141

1142 **Figure 8**



1143

1144 **Figure 8: Using PRED as a measure of similarity and separability for connectomic data**

1145 **a** Illustration of the group-database (class-vector) structure used for comparing the two  
 1146 datasets, FAFB and Hemibrain. Each bin represents the average strength of connections  
 1147 between the LHNs belonging to the group and a single glomerulus (Glom). High value of  
 1148 group-database PRED confirms stereotypy between FAFB and Hemibrain datasets for all 4  
 1149 levels of groupings of lateral horn neurons (LHNs). Each value in the violin represents a pair  
 1150 of groups within the specified hierarchy level (connectivity type:  $n = 496$  pairs of  
 1151 connectivity types, cell type:  $n = 378$  pairs of cell types, tract:  $n = 66$  pairs of tracts, region:  $n$   
 1152  $= 6$  pairs of regions). The calculations were performed over the antennal lobe glomerulus to  
 1153 LHN connectivity data. The connectivity values were averaged over all all neurons within the  
 1154 specified groups. **b** Illustration of the group-neuron (class-sample) dataset for calculating the  
 1155 across-group separability of neuron connectivity patterns. Each column contains the  
 1156 connectivity vectors of all LHNs belonging to a group. Each bin represents the strength of  
 1157 connections between an LHN and a single glomerulus. Group-neuron PRED for the dataset  
 1158 with individual neurons grouped into connectivity types (FAFB:  $n = 36$  pairs of connectivity  
 1159 types, Hemibrain:  $n = 276$ ), cell types (FAFB:  $n = 21$  pairs of cell types, Hemibrain:  $n = 210$ ),  
 1160 tracts (FAFB:  $n = 36$  pairs of regions, Hemibrain:  $n = 45$ ) or regions (FAFB:  $n = 6$  pairs of  
 1161 tracts, Hemibrain:  $n = 6$ ) for each of the two datasets, FAFB and Hemibrain. Black horizontal  
 1162 line represents the mean, and error bars represent s.e.m.

1163

1164 **Table 1**

Metric	Range	Chance Level	Discreteness	Consistency with global scaling	Consistency with global translation
PRED	[-1 1]	0	Continuous	Constant	Constant
PC	[-1 1]	0	Discrete for 2 classes	Constant	Constant
COS	[-1 1]	0	Continuous	Constant	Changes
MAN	[0 1]	$> 0^a$	Continuous	Changes	Constant
CHEB	[0 1]	$> 0^a$	Continuous	Changes	Constant
EUC	[0 1]	$> 0^a$	Continuous	Changes	Constant

1165 *Table 1: Summary of the properties of class-vector metrics*

1166 Values in red represent less desirable behavior compared with PRED.

1167 <sup>a</sup> Chance level of these metrics varies with the number of classes.

1168

1169 **Table 2**

1170

Metric	Range	Chance Level <sup>a</sup>	Discreteness	Consistency with number of classes	Consistency with number of samples	Dynamic range <sup>b</sup>	Variability <sup>b</sup>
PRED	[-1 1]	0	Continuous	Constant	Constant	-	-
ETM	[0 1]	0.5 <sup>c</sup>	Discrete	Changes	Changes	Smaller	Higher
SIL	[-1 1]	0 <sup>c</sup>	Continuous	Changes	Constant	Similar	Similar
DBI	[Inf 0]	>0 <sup>c</sup>	Continuous	Changes	Constant	Smaller	Higher
DUNN	[0 Inf]	>0 <sup>c</sup>	Continuous	Changes	Changes	Smaller	Higher
CH	[0 Inf]	>0 <sup>c</sup>	Continuous	Changes	Changes	Smaller	Higher

1171 **Table 2: Summary of the properties of class-sample metrics**

1172 Values in red represent less desirable behavior compared with PRED.

1173 <sup>a</sup> reported for a dataset with 2 classes

1174 <sup>b</sup> as compared to PRED

1175 <sup>c</sup> Chance level of these metrics varies with the number of classes.

1176