



This is a repository copy of *Compensatory variability in network parameters enhances memory performance in the Drosophila mushroom body.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/171061/>

Version: Submitted Version

Article:

Abdelrahman, N.Y. orcid.org/0000-0003-3198-3225, Vasilaki, E. orcid.org/0000-0003-3705-7070 and Lin, A.C. orcid.org/0000-0001-6310-9765 (Submitted: 2021) Compensatory variability in network parameters enhances memory performance in the Drosophila mushroom body. bioRxiv. (Submitted)

<https://doi.org/10.1101/2021.02.03.429444>

© 2021 The Authors. Article made available under a CC-BY-NC 4.0 International license (<http://creativecommons.org/licenses/by-nc/4.0/>).

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Compensatory variability in network parameters enhances memory performance in the *Drosophila* mushroom body

Nada Y. Abdelrahman^{1,3}, Eleni Vasilaki^{2,3}, and Andrew C. Lin^{1,3,*}

¹*Department of Biomedical Science, University of Sheffield, Sheffield, UK*

²*Department of Computer Science, University of Sheffield, Sheffield, UK*

³*Neuroscience Institute, University of Sheffield, Sheffield, UK*

**For correspondence: andrew.lin@sheffield.ac.uk*

February 3, 2021

Abstract

Neural circuits use homeostatic compensation to achieve consistent behaviour despite variability in underlying intrinsic and network parameters. However, it remains unclear how compensation regulates variability across a population of the same type of neurons within an individual, and what computational benefits might result from such compensation. We address these questions in the *Drosophila* mushroom body, the fly's olfactory memory center. In a computational model, we show that memory performance is degraded when the mushroom body's principal neurons, Kenyon cells (KCs), vary realistically in key parameters governing their excitability, because the resulting inter-KC variability in average activity levels makes odor representations less separable. However, memory performance is rescued while maintaining realistic variability if parameters compensate for each other to equalize KC average activity. Such compensation can be achieved through both activity-dependent and activity-independent mechanisms. Finally, we show that correlations predicted by our model's compensatory mechanisms appear in the *Drosophila* hemibrain connectome. These findings reveal compensatory variability in the mushroom body and describe its computational benefits for associative memory.

Significance statement

How does variability between neurons affect neural circuit function? How might neurons behave similarly despite having different underlying features? We addressed these questions in neurons called Kenyon cells, which store olfactory memories in flies. Kenyon cells differ among themselves in key features that affect how active they are,

26 and in a model of the fly’s memory circuit, adding this inter-neuronal variability made the model fly worse at
27 learning the values of multiple odors. However, memory performance was rescued if compensation between the
28 variable underlying features allowed Kenyon cells to be equally active on average, and we found the hypothesized
29 compensatory variability in real Kenyon cells’ anatomy. This work reveals the existence and computational benefits
30 of compensatory variability in neural networks.

31 **Introduction**

32 Noise and variability are inevitable features of biological systems. Neural circuits achieve consistent activity pat-
33 terns despite this variability using homeostatic plasticity: because neural activity is governed by multiple intrinsic
34 and network parameters, variability in one parameter can compensate for variability in another to achieve the
35 same circuit behaviour [1–5]. This phenomenon of compensatory variability has typically been addressed from the
36 perspective of consistency of neural activity across individual animals [6, 7] or over an animal’s lifetime, in the face
37 of circuit perturbations [8–11]. However, less attention has been paid to potential benefits of maintaining consistent
38 neuronal properties across a population of neurons within an individual circuit.

39 Indeed, previous work has emphasized the benefits of neuronal heterogeneity rather than neuronal homogeneity
40 [12–14]. Of course, different neuronal classes encode different information (e.g., visual vs. auditory neurons, or ON
41 vs. OFF cells). Yet even in populations that ostensibly encode the same kind of stimulus, like olfactory mitral cells,
42 heterogeneity of neuronal excitability can increase the information content of their population activity [15–17]. In
43 addition, heterogeneity in neuronal time scales can improve learning in neural networks [18, 19]. In what contexts
44 and in what senses might the opposite be true, i.e., when does neuronal similarity provide computational benefits
45 over neuronal diversity? And what mechanisms could enforce neuronal similarity in the face of inter-neuronal
46 variability?

47 Here we address these questions using olfactory associative memory in the mushroom body of the fruit fly
48 *Drosophila*. Flies learn to associate specific odors with salient events (e.g., food or danger). These olfactory
49 associative memories are stored in the principal neurons of the mushroom body, called Kenyon cells (KCs), as
50 modifications in KCs’ output synapses [20–22] (reviewed in [23]). Because learning occurs at the single output
51 layer, the nature of the odor representation in the KC population is crucial to the fly’s ability to learn to form
52 distinct associative memories for different odors. In particular, the fact that KCs respond sparsely to incoming
53 odors ($\approx 10\%$ per odor) [24] allows different odors to activate unique, non-overlapping subsets of KCs and thereby
54 enhances flies’ learned discrimination of similar odors [25].

55 A potential problem for this sparse coding arises from variability between KCs. KCs receive inputs from
56 second-order olfactory neurons called projection neurons (PNs), with an average of ≈ 6 PN inputs per KC, and
57 typically require simultaneous activation of multiple input channels in order to spike [26], thanks to high spiking
58 thresholds and feedback inhibition [25, 27]. However, there is substantial variation across KCs in the key parameters
59 controlling their activity, such as the number of PN inputs per KC [28], the strength of PN-KC synapses, and KC

60 spiking thresholds [27]. Intuitively, such variation could lead to a situation where some KCs with low spiking
61 thresholds and many or strong excitatory inputs fire indiscriminately to many different odors, while other KCs
62 with high spiking thresholds and few or weak excitatory inputs never fire; KCs at both extremes are effectively
63 useless for learning to classify odors, even if overall only 10% of KCs respond to each odor. However, it remains
64 unclear whether biologically realistic inter-KC variability would affect the mushroom body’s memory performance,
65 and what potential strategies might counter the effects of inter-KC variability.

66 Here we show in a rate-coding model of the mushroom body that introducing experimentally-derived inter-KC
67 variability into the model substantially impairs its memory performance. This impairment arises from decreased
68 dimensionality of the KC population activity and increased similarity between KC responses to different odors,
69 ultimately arising from the variability in average activity among KCs. However, memory performance can be
70 rescued by compensating away variability in KC activity while preserving the experimentally observed variation in
71 the underlying parameters. This can occur through activity-dependent homeostatic plasticity or direct correlations
72 between key parameters like number vs. strength of inputs. Finally, we analyze the hemibrain connectome to show
73 that indeed, the number of PN inputs per KC is inversely correlated with the strength of each input, while the
74 strength of inhibitory inputs is correlated with the total strength of excitatory inputs. Thus, we show both the
75 existence and computational benefit of compensatory variability in mushroom body network parameters.

76 Results

77 Realistic inter-KC variability impairs memory performance

78 To study how variability between KCs might affect the fly’s olfactory memory performance, we modelled the
mushroom body as a rate-coding neural network (Fig. 1). To simulate the input activity from PNs, we modeled

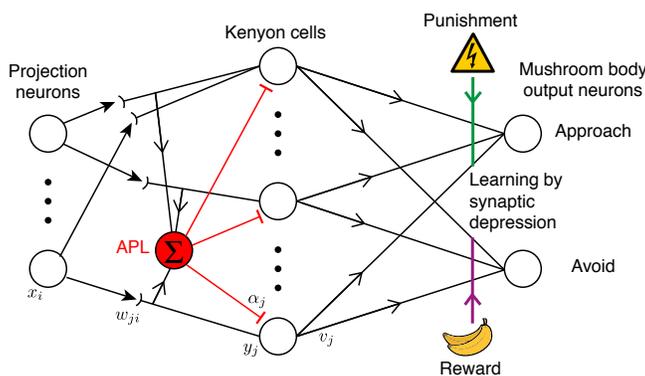


Figure 1: Schematic for the mushroom body network model. Projection neurons in the input layer relay the odor responses, x_i , downstream to the Kenyon cells (y_j). Kenyon cells connect randomly to the projection neurons with synaptic weights w_{ji} and receive global inhibition from the APL neuron with weight α_j . Learning occurs when dopaminergic neurons (DANs) carrying punishment (reward) signals from the environment depress the synapses (v_j) between the active Kenyon cells and the mushroom body output neurons (MBONs) that lead to approach (avoidance) behavior.

79

80 their activity as a saturating non-linear function of activity of the first-order olfactory receptor neurons (ORNs)

81 (see Methods; [29]). We applied this function to the recorded odor responses of 24 different olfactory receptors [30]
82 to yield simulated PN activity, as has been done in many computational studies of fly olfaction [31–34]. To simulate
83 variability in PN activity across different encounters with the same odor, we created several ‘trials’ of each odor
84 and added Gaussian noise to PN activity, following the coefficients of variation reported in [35]. To increase the
85 number of stimuli beyond the 110 recorded odors in [30], we generated odor responses in which the activity of each
86 PN was randomly sampled from that PN’s activity across the 110 odors used in [30] (results were similar with the
87 ‘real’ 110 odors; see Methods and below).

88 Each KC in our model received excitatory input from a randomly selected set of N PNs, each with strength w .
89 A KC’s response to each odor was the sum of excitatory inputs minus inhibition, minus a spiking threshold θ ; if
90 net excitation was below the threshold, the activity was set to zero. Inhibition came from the feedback interneuron
91 APL (‘Anterior Paired Lateral’), which is excited by and inhibits all KCs [25]. To avoid simulating the network
92 in time, we simplified the feedback inhibition into pseudo-feedforward inhibition, in which APL’s activity was the
93 sum of all post-synaptic excitation of all KCs (without the KCs’ threshold applied); we based this simplification on
94 the fact that KCs and APL form reciprocal synapses with each other on KC dendrites (i.e., before the KCs’ spike
95 initiation zone), and APL activity is somewhat spatially restricted between KC axons and dendrites [36].

96 Learning in flies occurs when KCs (responding to odor) are active at the same time as dopaminergic neurons
97 (DANs, responding to ‘reward’ or ‘punishment’); the coincident activity modifies the output synapse from KCs onto
98 mushroom body output neurons (MBONs) that lead to behavior (e.g., approaching or avoiding an odor). Typically,
99 the output to the ‘wrong’ behavior is depressed: for example, pairing an odor with electric shock weakens the output
100 synapses from KCs activated by that odor onto MBONs that lead to ‘approach’ behavior [21, 22, 37, 38] (reviewed
101 in [23]). We simulated this plasticity using a simplified architecture with only two MBONs, one for ‘approach’ and
102 one for ‘avoid’. The input odors were randomly divided: half were paired with punishment and half with reward.
103 During training, KCs activated by rewarded odors weakened their synapses onto the ‘avoid’ MBON, while KCs
104 activated by punished odors weakened their synapses onto the ‘approach’ MBON (depression by exponential decay;
105 see Methods). The fly’s behavior then depended probabilistically (via a softmax function; see Eq. 21, Methods)
106 on whether the ‘avoid’ or ‘approach’ MBON’s was greater, and the model’s accuracy in learning was scored as
107 the fraction of correct decisions for unseen noisy variants of the trained odors (i.e., avoiding punished odors and
108 approaching rewarded odors).

109 To test the effect of realistic inter-KC variability on this model, we introduced variability step-by-step. We first
110 tested the performance of the model holding constant across all KCs the 3 parameters N (number of PN inputs per
111 KC), w (strength of each PN-KC connection) and θ (KC spiking threshold). Then we added inter-KC variability
112 step-by-step: first varying only one out of 3 parameters, then 2 out of 3, then all 3 parameters (thus 8 possible
113 models). Inter-KC variability in N , w and θ followed experimentally measured distributions (Fig. 2A1-3) [27, 28].
114 Increasing inter-KC variability systematically degraded the model’s performance when tested on 100 input odors:
115 the more variable parameters there were, the worse the performance (Fig. 2B). In the two extreme cases, the model

116 with all 3 parameters fixed performed at 72.5% accuracy while the model with all 3 parameters variable performed
 117 at 64% accuracy. This performance trend was the same when these 8 models were trained and tested on the real
 118 input odors responses from [30] (78.1% v. 63.9% Fig. S1).

119 To test whether this effect is robust to different learning and testing conditions, we tested the two extreme
 120 cases while varying the numbers of input odors to be classified, the amount of noise in PN activity, the learn-
 121 ing rate at the KC-MBON synapse (the two models might have different optimal learning rates: η in Eq. 20),
 122 or the indeterminacy of the fly's decision making (c in the softmax equation, Eq. 21). In every case, the model
 123 with all parameters fixed (which we call the 'homogeneous' model) consistently outperformed the model with all
 124 parameters variable (which we call the 'random' model) (Fig. 2C1-4). These results indicate that biologically re-
 125 alistic variability in KC network parameters impairs the network's ability to classify odors as rewarded vs. punished.

126

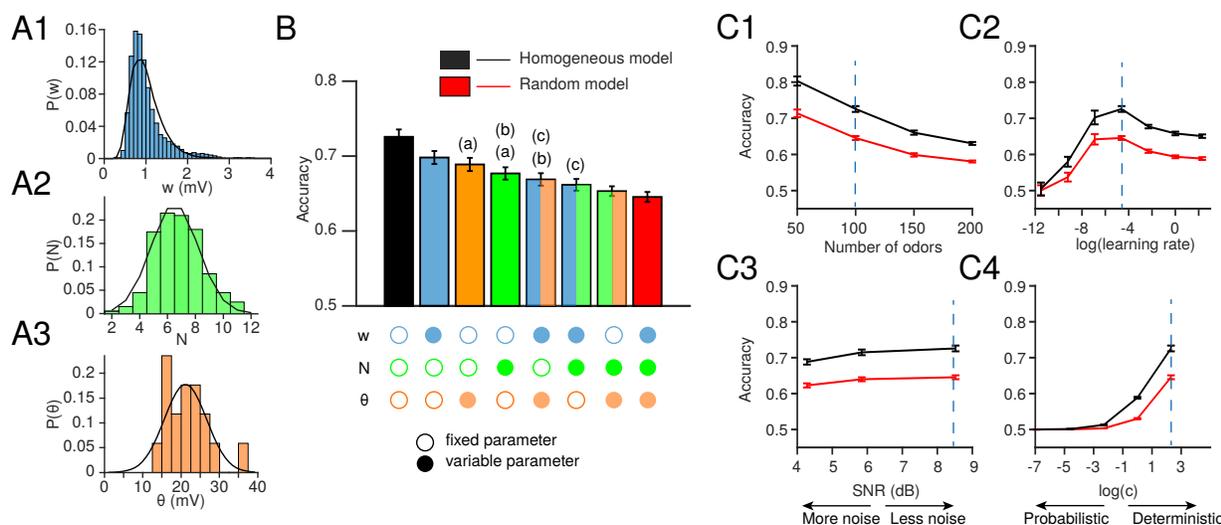


Figure 2: Inter-KC variability in w , N and θ degrades the model fly's memory performance.

(A) Histograms of the experimentally measured distributions for: (A1) w (amplitude of spontaneous excitatory postsynaptic potentials in KCs, mV; data from [27]), (A2) N (number of PN inputs per KC, measured as the number of dendritic 'claws'; data from [28]), (A3) θ (spiking threshold minus resting potential, mV; data from [27]). The overlaid black curves show log-normal (w) and Gaussian (N , θ) fits to the data.

(B) The model fly's memory performance (given 100 input odors), varying the parameters step by step. Fixed and variable parameters are shown by empty and filled circles, respectively. The homogeneous model (all parameters fixed; black) performs the best and the random model (all parameters variable; red) performs the worst. All bars are significantly different from each other unless they share the same letter annotations (a, b, etc.), $p < 0.05$ by Wilcoxon signed-rank test (for matched models with the same PN-KC connectivity) or Mann-Whitney test (for unmatched models with different PN-KC connectivity, i.e., fixed vs. variable N), with Holm-Bonferroni correction for multiple comparisons (full statistics in Table S1). $n = 25$ model instances with different random PN-KC connectivity, error bars show twice the standard error of the mean (95.4% confidence interval).

(C) The performance trend is consistent over a range of different conditions: (C1) number of input odors, (C2) the learning rate used to learn the optimum weights between KCs and MBONs, (C3) amount of noise in PN activity (measured by signal-to-noise ratio, SNR), (C4) the indeterminacy in the decision making, quantified by $\log(c)$, where c is the constant in the soft-max function (Eq. 21). The vertical dotted lines indicate the conditions used in panel B.

127 Realistic inter-KC variability reduces separation between KC odors representations

128 We next asked what features of the KC population odor representation might account for the worse performance
129 of the random model compared to the homogeneous model. Learning the optimal KC-MBON weights to correctly
130 classify the rewarded versus punished odors is equivalent to finding a hyper-plane (in 2000-dimensional space) to
131 separate KC responses to rewarded odors from those to punished odors. Therefore, a model with better separability
132 between KC odor representations would find a better separating hyper-plane, and have better performance in
133 classifying unseen noisy variants of the trained odors. We measured separability using a variety of metrics.

134 We first asked whether odors are more widely separated in KC coding space in the homogeneous model, using
135 angular distance, a scale-insensitive distance metric (see Methods). For each odor, we took the centroid of KC
136 responses to the noisy variants of that odor, and for each pair of odors, we measured the angular distance between
137 their respective centroids (Fig. 3A1). Indeed, the angular distance between odors (averaged across all odor pairs)
138 was larger in the homogeneous model (Fig. 3A2), which matched the higher accuracy (Fig. 3A3, where each dot
139 represents one instantiation of the network). This difference also extended to the angular distance between the
140 centroids of the groups of odors randomly assigned to be rewarded and punished (Fig. 3B2), suggesting that the
141 greater inter-odor distances in the homogeneous model make it easier to draw a hyper-plane separating the rewarded
142 and punished odors.

143 However, the separability of clusters of noisy variants of odors might depend not only on the distance between
144 their centroids, but also on their variability. For instance, two clusters of noisy variants with well separated centroids
145 might overlap if the data points in the clusters are not tightly packed. Therefore, we next measured the quality of
146 clustering in each model using the Davies-Bouldin Index (DBI). DBI measures the variance within clusters divided
147 by the distance between the centroids of each cluster [39], so high DBI means more overlapping, less separable
148 clusters. When we calculated DBI using different pairs of odors (Fig. 3C1), treating each odor (with its noisy
149 variants) as its own cluster, DBI values were similar in the random and homogeneous models (Fig. 3C2), suggesting
150 that poor performance in the random model was not explained by poor clustering of noisy variants (Fig. 3C3). (The
151 DBI was slightly higher in the random model using the original odors from [30]: Fig. S1). However, DBI was higher
152 in the random model when considering the two clusters of all rewarded odors vs. all punished odors (Fig. 3D1-2),
153 and showed a weak inverse correlation with memory performance (Fig. 3D3) (note that each instantiation of the
154 network received the same odors but different random reward/punishment assignments). These results suggest that
155 in the homogeneous model (compared to the random model), odor representations are arranged in KC coding space
156 in a way to allow punished and rewarded odors to be more easily separated.

157 We hypothesized that odor responses in the homogeneous model are more separable because they are arranged
158 across more dimensions in KC coding space, allowing them more degrees of freedom. We quantified dimensionality
159 according to [40]. Dimensionality of a dynamic system is the number of independent dimensions that define the
160 system's response to a given input. For example, if a system nominally has 3 dimensions but all its responses lie on
161 a straight line, its dimensionality is only 1, in contrast to a system whose responses are distributed throughout the

162 3-dimensional space (Fig. 3E1). We found that KC responses in the homogeneous model had a significantly higher
 163 dimensionality than those in the random model (Fig. 3E2), matching the higher performance in the homogeneous
 164 model (Fig. 3E3). Together, these metrics indicate that introducing the realistic inter-KC variability in w , N ,
 165 and θ worsens the performance of the network by reducing the dimensionality (and thus separability) of KC odor
 166 representations.

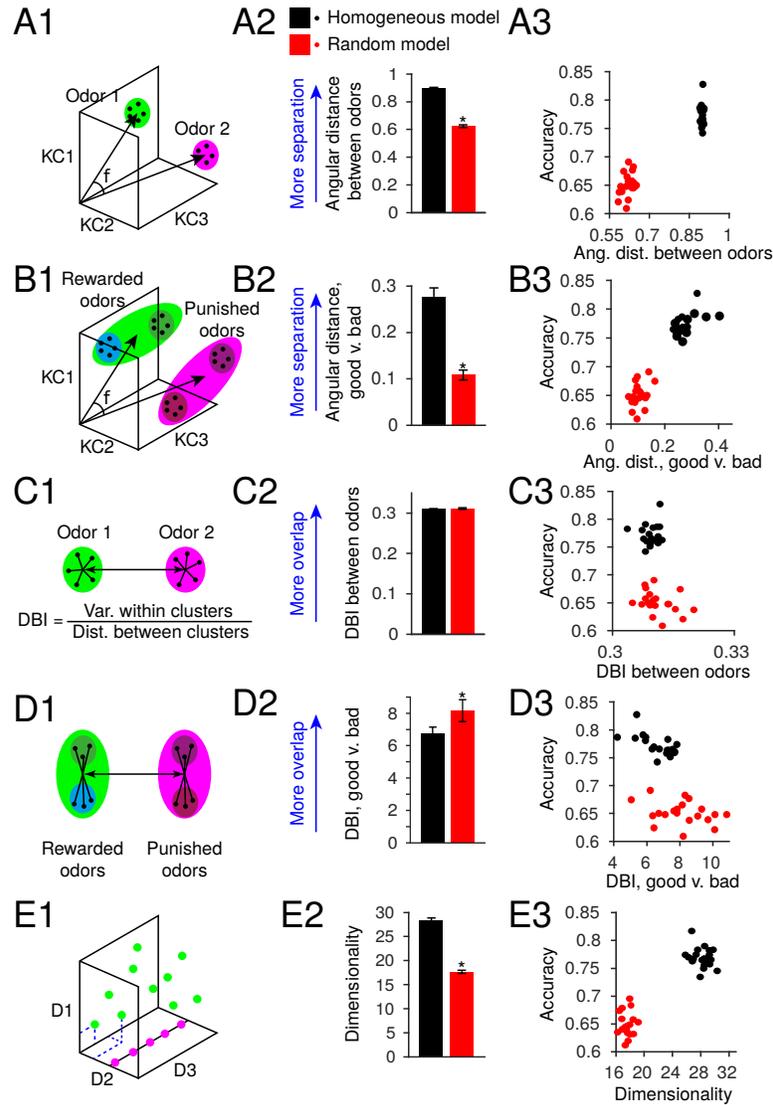


Figure 3: Inter-KC variability in w , N and θ reduces separability of KC odor representations.

Left column: (A1-D1) show schematic illustrations of separability metrics: angular distance between individual odors (A1) or between rewarded vs. punished odors (B1), and DBI between individual odors (C1) or between rewarded vs. punished odors (D1). (E1) shows the dimensionality of a system with 3 variables. The system with its states scattered throughout 3D space (green) has dimensionality 3 while the system with all states on a single line (magenta) has dimensionality 1.

Middle column: Separability metrics for the homogeneous (black) and random (red) models. Compared to the random model, the homogeneous model has higher angular distance (A2,B2), similar DBI between odors (D2) but lower DBI between rewarded vs. punished odors (D2), and higher dimensionality (E2). Error bars show twice the standard error of the mean. * difference between homogeneous and random models, $p < 0.05$, Mann-Whitney test (full statistics in Table S1).

Right column: (A3-E3) Scatter plots show performance vs. separability metric in the respective rows, calculated in $n=25$ random instantiations of the network.

167 Realistic inter-KC variability weakens specialization of KC responsiveness

168 We hypothesized that the lower dimensionality of the random model might arise because fewer KCs provide useful
169 odor identity information when some are indiscriminately active while others are completely silent. Sparse coding
170 requires sparseness in two dimensions: population sparseness (each stimulus activates few neurons) and lifetime
171 sparseness (each neuron responds to few stimuli) [41]. While our models enforced population sparseness by scaling
172 inhibition and spiking thresholds to achieve a coding level (fraction of cells active per odor) of 0.1 (averaged across
173 all odors), they did not enforce any particular lifetime sparseness. In an extreme case, a model could have very
174 consistent population sparseness with a coding level of 0.1 for all odors, simply by having the same 10% of cells
175 responding to every odor and the other 90% being completely silent. In this case, none of the cells would provide
176 any useful information about odor identity and dimensionality would be 0. We asked whether a less extreme version
177 of this problem could explain the lower dimensionality and memory performance of the random model.

178 We measured the lifetime sparseness of KCs in the homogeneous and random models. Lifetime sparseness
179 quantifies how specialized a cell is to particular input stimuli: 1 means a cell fires to one stimulus and no other
180 stimuli, while 0 means it fires equally to all stimuli. A cell that fires to no stimuli has an undefined sparseness (see
181 Methods). The homogeneous model had fairly consistent lifetime sparseness values, with almost 90% of KCs having
182 a lifetime sparseness between ~ 0.85 and 1. In contrast, the random model had KCs with much more variable lifetime
183 sparseness, with a long tail of KCs with low sparseness (below 0.7) and more than 40% of KCs having undefined
184 sparseness (i.e., completely silent). The contrasting distributions of lifetime sparseness can be seen in the cumulative
185 distribution functions (cdfs) of lifetime sparseness in Fig. 4A, in how the steep curve of the homogeneous model
186 and the shallow curve of the random model cross each other. This result can also be seen in the larger standard
187 deviation of lifetime sparseness across KCs in the random model (Fig. 4B). The silent KCs can be seen as the
188 fraction of missing KCs needed for the cdf curves to reach 1; the random model has many more silent KCs than
189 the homogeneous model (Fig. 4A). Because silent KCs are useless for odor identity coding, a high number of silent
190 KCs corresponds to low dimensionality of KC odor representations (Fig. 4C).

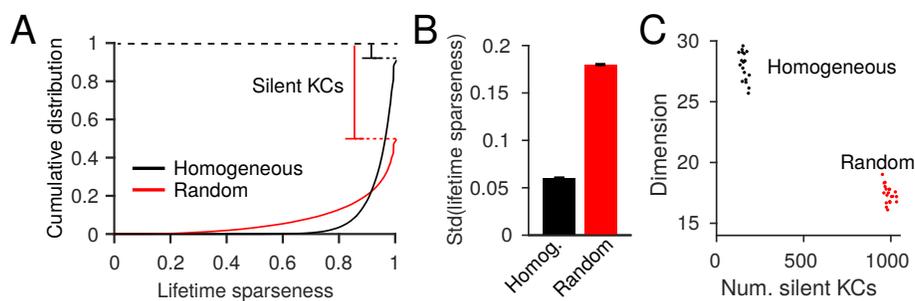


Figure 4: Inter-KC variability increases variability of lifetime sparseness and fraction of silent KCs
(A) Cumulative distribution function (cdf) of the lifetime sparseness of KCs in the homogeneous (black) and random (red)
models, across 20 instantiations of the network. The gap between 1.0 and the top of the cdf represents silent KCs (lifetime
sparseness undefined). (B) The random model has larger standard deviation in lifetime sparseness among KCs. Error bars
show twice the SEM, $n = 20$ random instantiations of the network. Bars are different, $p < 0.05$, Mann-Whitney test (see
Table S1). (C) Number of silent KCs plotted versus the dimensionality of KCs; each dot is one random model instance.

191 **Compensatory tuning of KC parameters rescues memory performance**

192 Because the central problem for memory performance in the random model was inter-KC variability in average
193 levels of activity, we hypothesized that performance could be rescued in models where KCs could achieve roughly
194 equal activity across the population, while still preserving experimentally realistic variability in spiking thresholds
195 and number/strength of excitatory inputs.

196 **Parametric tuning of excitatory input weights**

197 First, we tested a model that equalizes KC activity indirectly, by making parameters compensate for each other in
198 an activity-independent way. In particular, we modeled KCs as adjusting input synaptic weights (w) to compensate
199 for variability in spiking threshold (θ) and number of PN inputs (N). Thus, an individual KC with low θ or high
200 N would have low w , while a KC with high θ or low N would have high w . We simulated these correlations
201 ($w \propto \sqrt{\theta}$; $w \propto 1/\sqrt{N}$) constrained by experimental data. To do this, we sampled N and θ from the distributions
202 in Fig. 2A, and sampled w from a posterior compensatory distribution, $P(w | n, \theta)$, whose overall shape across all
203 KCs was constrained to be the same as the experimental $P(w)$ in Fig. 2A1 but which was composed of multiple
204 distributions of $P(w)$ for different values of N and θ . For example, a KC with a relatively high $n = 7$ would sample
205 its weights from a $P(w)$ shifted to the left (lower w) (Fig. 5A1, dashed lines), while a KC with a relatively high
206 $n = 2$ would sample its weights from a $P(w)$ shifted to the right (higher w) (Fig. 5A1, solid lines). The same would
207 be true for different values of θ (Fig. 5A1, different shadings). We fitted these component $P(w)$ curves so that with
208 experimentally observed distributions of N and θ , the sum of the components would produce the experimentally
209 observed distribution of w across all KCs (see Methods). (Note that this algorithm is not meant to describe an
210 actual biological mechanism, merely to create correlations between w vs. N and θ while constraining the parameters
211 to experimentally realistic distributions. Biologically, such correlations could arise through several mechanisms; see
212 Discussion.) This compensatory mechanism rescued the fly's performance, producing significantly higher accuracy
213 at classifying odors than the random model (Fig. 5B, cyan bars), likely resulting from the higher dimensionality of
214 KC representations (Fig. 5C) and reduced variability in KC lifetime sparseness (Fig. 5D).

215 **Activity-dependent tuning of KC parameters**

216 We next tested compensatory mechanisms based on activity rather than explicit correlations between network
217 parameters. Here, each KC has the same desired average activity level across all odors, A_0 (with a tolerance of
218 $\pm 6\%$). We tested three models, each of which equalized average KC activity A_0 by tuning a different parameter:
219 input excitatory weights (w), inhibitory weights (α), or spiking thresholds (θ). The non-tuned parameters followed
220 the distributions in Fig. 2A (inhibitory weights were constant when non-tuned), while individual KCs adjusted the
221 tuned parameter according to whether their activity was too high or too low. For example, a relatively highly active
222 KC (whether because it has high w or N , low θ , or simply receives input from highly active PNs) would scale down
223 its excitatory weights (Fig. 5A2), scale up its inhibitory weights (Fig. 5A3), or scale up its spiking threshold (Fig.

224 5A4). Likewise, a relatively inactive (or indeed totally silent) KC would do the reverse (see Methods for details of
225 the update rules underlying the homeostatic tuning and discussion of variant update rules shown in Figs. S3,S4).

226 All three homeostatic models performed as well as the homogeneous model (Fig. 5B1, blue, green, magenta bars),
227 and indeed even out-performed the homogeneous model when decision-making was more stochastic (lower value of
228 c in the softmax function; Fig. 5B2). The more stochastic decision-making makes the task more difficult and thus
229 brings out the enhanced coding by the homeostatic models. Indeed, the dimensionality of KC odor representations
230 in the homeostatic models was even higher than that in the homogeneous model (Fig. 5C), and the variability in
231 KC lifetime sparseness was even lower (Fig. 5D).

232 What distributions of excitatory weights, inhibitory weights, or spiking thresholds emerge after activity-dependent
233 tuning to equalize KC activity? Do they match experimentally observed distributions? Tuning excitatory weights
234 led to a distribution fairly similar to the approximately log-normal experimentally observed distribution of EPSP
235 amplitudes (Fig. 5E). Tuning spiking thresholds led to a distribution with greater variance than the experimental
236 distribution, although with a qualitatively similar Gaussian shape (Fig. 5F). This larger variance of thresholds
237 suggests that natural variation of θ is too small, on its own, to equalize KC activity given the variation in the
238 number/strength of excitatory inputs.

239 The tuned distribution of inhibitory weights differed even more strongly from experimental results. While there
240 are no experimental measurements of inhibitory weights, equalizing KC activity by tuning inhibitory weights re-
241 quired many of them to be negative (Fig. 5G), which is unrealistic, because negative inhibition is actually excitation,
242 and there are no reports of GABAergic excitation of KCs [42].

243 Why did our model require negative inhibition? This result can be understood by considering one of the model's
244 constraints: that inhibition is only strong enough to reduce the fraction of active KCs by half, i.e., 10% of KCs are
245 active on average in normal flies, while 20% of KCs are active if inhibition is blocked (based on results from [25]).
246 Because the average threshold must be high enough that 80% of KCs are silent on average even without inhibition,
247 the wide variation in thresholds and excitation means that many KCs will have excitation so weak, and thresholds
248 so high, that no stimulus could ever drive them above threshold, even in the absence of inhibition. For inhibition
249 to compensate for inactivity even in the absence of inhibition, it must become negative (i.e., excitatory) in these
250 weakly-activated KCs. In contrast, the models that tune excitatory weights or thresholds do not face this problem,
251 because inactive KCs can simply increase their excitatory weights or decrease their thresholds. The central problem
252 for the inhibitory plasticity model is that inhibition is not a strong enough force in our system to balance out
253 variable excitation and thresholds on its own without becoming negative. Indeed, if we relax the constraint that
254 coding level be 0.2 without inhibition, such that sparseness is enforced by inhibition alone (not thresholds), then
255 variable inhibition can equalize KC activity without becoming negative (Fig. 5G). However, in this case, the coding
256 level without inhibition was 99.7% (Fig.5G), which is not observed experimentally [25]. Even allowing a coding
257 level without inhibition of 50%, equalizing KC activity still requires some APL-KC inputs to be negative (Fig. 5G).
258 Overall, these results suggest that tuning inhibitory weights cannot compensate on its own for variability in other

259 KC parameters. More likely, the system optimizes multiple parameters at once (see Fig. 7 and Discussion).

260 To better understand why equalizing average activity improves performance, we asked whether memory perfor-
 261 mance can also be rescued by equalizing not KC average activity, but rather KC response probability (equivalent
 262 to average activity if KC activity is binarized, i.e., 0 or 1). Equalizing response probability (as opposed to average
 263 activity) by tuning KC spiking thresholds has been shown to improve separation of KC odor representations in a dif-
 264 ferent computational model [34]. However, in our model, this technique (tuning thresholds to equalize KC response
 265 probability) produced worse classification performance and lower dimensionality compared to tuning thresholds to
 266 equalize KC average activity (Fig. S4A,B), though still better than the random model (compare Fig. S4 to Fig. 5).
 267 This result can be understood by considering that dimensionality of neuronal activity is maximized when variance
 268 along all dimensions is equal (Fig. 3) [40], but equalizing KC response probability still allows KCs to have unequal
 269 average activity (one KC's supra-threshold activity might be higher than another's), which would cause KCs to
 270 differ between each other in their variances in activity across odors (a KC's variance in activity depends on its
 271 average activity because its response to most odors is zero).

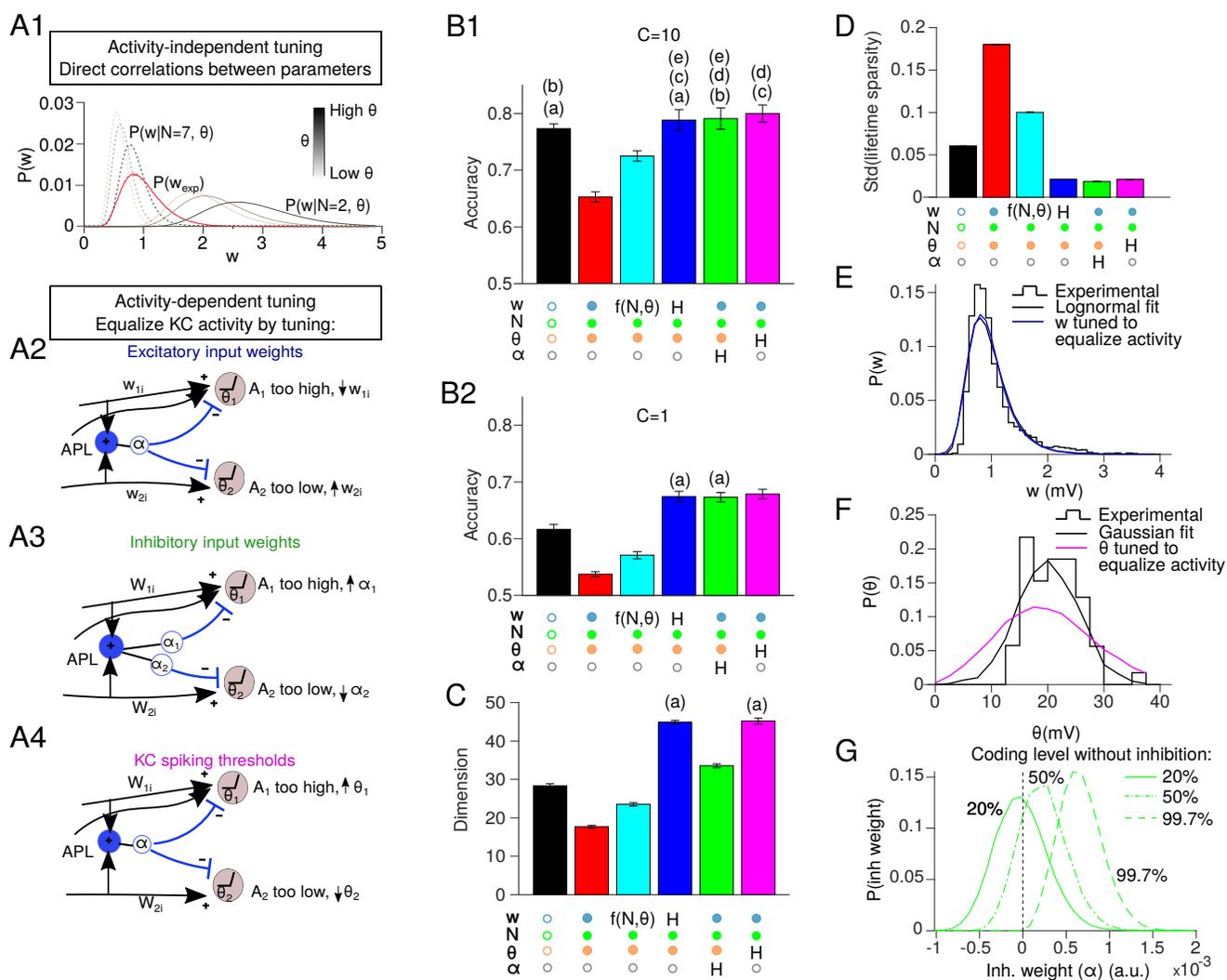


Figure 5: Compensation in network parameters rescues memory performance.

(A) Schematics of different compensation methods. **(A1)** Lognormal fit of experimental distribution of the synaptic weights $P(w_{exp})$ (red), and its component distributions for different w and θ , $P(w | N, \theta)$, for high $N = 7$ (dotted) or low $N = 2$ (solid). Shadings of gray indicate different values of θ . **(A2-4)** Mechanisms for activity-dependent homeostatic compensation. Overly active KCs weaken excitatory input weights (w_{ji} , A2), strengthen inhibitory input weights (α_j , A3), or raise spiking thresholds (θ_j , A4). Inactive KCs do the reverse.

(B1) Compensation rescues performance, alleviating the defect caused by inter-KC variability in the random model (red) compared to the homogeneous model (black), whether compensation occurs by setting w according to N and θ (cyan; A1), using activity-dependent homeostatic compensation to adjust excitatory weights (dark blue; A2), inhibitory weights (green; A3) or spiking thresholds (magenta; A4). **(B2)** Differences between models are more apparent when the task is more difficult due to more stochastic decision-making ($c = 1$ instead of $c = 10$ in the softmax function in Eq. 21).

(C-D) Dimensionality of KC representations (C) and standard deviation of KC lifetime sparseness (D) in the models described above. Activity-dependent models have the highest dimensionality and lowest variability in KC sparseness.

$n = 20$ model instances with different random PN-KC connectivity. Error bars show two times the SEM, i.e., 95.4% confidence interval. All bars are significantly different from each other unless they share the same letter annotations, $p < 0.05$, by Wilcoxon signed-rank test (for matched models with the same PN-KC connectivity) or Mann-Whitney test (for unmatched models with different PN-KC connectivity, i.e., fixed vs. variable N), with Holm-Bonferroni correction for multiple comparisons (full statistics in Table S1). Annotations below bars indicate whether parameters were fixed (empty circle), variable (filled circle), or variable following a compensation rule ('H' for homeostatic tuning, $f(N, \theta)$ for parametric tuning).

(E) KC excitatory input synaptic weights (w) after tuning to equalize average activity (blue) follow a similar distribution to experimental data (black, from Fig. 2A1)

(F) KC spiking thresholds (θ) after tuning to equalize average activity (magenta) have wider variability than the experimental distribution (black, from Fig. 2A3).

(G) Tuning KC inhibitory weights (α) to equalize average activity requires many inhibitory weights to be negative, unless the coding level without inhibition is as high as 99.7%.

272

273

274 Robustness of pre-tuned compensations in new environments with novel odors

275 Any activity-dependent tuning depends on the model's context. If a fly tunes its network parameters based on
276 experience in one odor context (e.g., smelling only odors of one chemical family), will it still perform well at
277 classifying odors in a novel environment with different odors (e.g., odors of a different chemical family)? We
278 hypothesized that performance would depend more on tuning context with the activity-dependent compensation
279 mechanisms than the activity-independent mechanism where input weights were picked depending on N and θ
280 rather than activity.

281 To test this, we tuned the parameters in our models using only a subset of odors from [30], grouped by chemical
282 class, and then trained and tested the models on odor-reward/punishment associations using the other odors. We
283 took the four chemical classes that had the most odors in the dataset: acids, terpenes, alcohols and esters. For each
284 class, we tuned the model's parameters on that class and then trained the model to classify odors in the other 3 classes
285 ('novel' environment). For matched controls, we trained models that had been tuned on the same 3 classes used
286 for training/testing ('familiar' environment). As expected, the three activity-dependent models performed worse
287 in novel environments than familiar environments, while the activity-independent model performed consistently
288 regardless of tuning environment (blue, green and magenta vs. cyan in Fig. 6C). However, in general, tuning odors
289 on one class but training/testing on different classes does not fatally damage the activity-dependent compensation
290 strategies: although performance is worse in novel environments, it remains better than the random model. Thus,
291 activity-dependent compensation is still a good strategy to overcome the pernicious effects of inter-KC variation,
292 even if the compensation environment differs from the classification environment (at least within the range of the

293 odors in [30]).

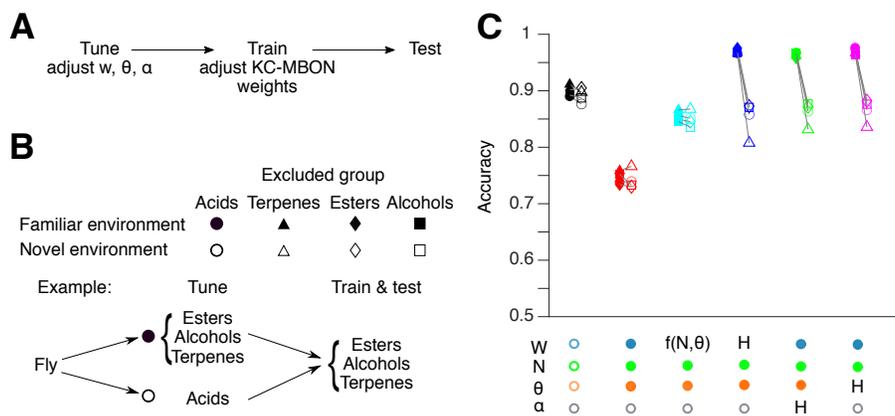


Figure 6: Robustness of pre-tuned compensations with novel odors.

(A) For each model fly, network parameters are tuned as in Fig. 5, on a subset of odors. At this stage, no rewards or punishments are given, and KC output weights are not modified. Then, the model is trained to classify rewarded and punished odors that are the same as or different from the odors used for tuning. Finally, the model is tested on new noisy variants of the odors used for training.

(B) Empty symbols (‘novel’ environment): models were tuned on odors from one chemical group (G_i : acids - circles, terpenes - triangles, esters - diamonds, or alcohols - squares), then trained and tested on odors from the other three groups ($G_{i \neq j}$). Each empty symbol is paired with a matched control (filled symbols) showing how that model would have fared in a ‘familiar’ environment: a model tuned, trained, and tested all on the same three groups of odors as the matched ‘novel’ model was trained and tested on ($G_{i \neq j}$).

(C) Models with activity-dependent compensation (blue, magenta, green) performed worse in novel environment than familiar environments (matching indicated by connecting lines). In contrast, models with no compensation (black, red), or compensation based on network parameters alone rather than activity (cyan), performed similarly in novel and familiar environments. Mean of 20 model instantiations, where each instantiation received a different permutation of odors (see Methods). Annotations below graph indicate whether parameters were fixed (empty circle), variable (filled circle), or variable following a compensation rule (‘H’ for homeostatic tuning, $f(N, \theta)$ for parametric tuning). Differences between novel and familiar environments, $p < 0.05$, Wilcoxon signed-rank test, except for: homogeneous model (black), esters; compensation by parametric tuning (cyan), acids, terpenes, esters (full statistics in Table S1).

294

295

296 Connectome reveals compensatory variation of input strength and numbers

297 Our proposed compensatory mechanisms predict correlations between the key model parameters. Excitatory weights
 298 (w) should be inversely correlated to number of PNs per KC (N) where w is tuned to compensate for variable N
 299 and θ (Fig. 7B) or where w is tuned to equalize KC activity (Fig. 7C). Meanwhile, inhibitory weights (α) should
 300 be positively correlated to the sum of excitatory weights ($\sum w$, or $\bar{w}N$, where \bar{w} is the mean w per KC) where
 301 inhibitory weights are tuned to equalize KC activity (Fig. 7D). Such correlations have been observed in larvae [43],
 302 but they have not yet been analyzed in the adult mushroom body.

303 To test these predictions, we analyzed the recently published hemibrain connectome [44, 45], which annotates
 304 all synapses between PNs and KCs in the right mushroom body of one fly. The connectome reveals three of our
 305 parameters: the number of PN inputs per KC (N), the strength of each PN-KC connection (w), and the strength
 306 of inhibitory inputs (α). Although the anatomy does not directly reveal w and α (which can only be measured

307 electrophysiologically), we used an indirect proxy for synaptic strength: the number of synapses per connection
308 (i.e., number of sites between two neurons where neuron 1 has a T-bar and neuron 2 has a postsynaptic density,
309 counted by machine vision; Fig. 7A). It seems reasonable to presume that, all else being equal, connections with
310 more synapses are stronger. Indeed, in the *Drosophila* antennal lobe, when comparing connections from ORNs to
311 ipsilateral PNs vs. contralateral PNs, ipsilateral connections are both stronger [46] and have more synapses per
312 connection [47]. Moreover, synaptic counts approximate synaptic contact area throughout the larval *Drosophila*
313 nervous system [48] and synaptic area approximates EPSP amplitude in mammalian cortex [49].

314 Therefore, to test if mean w and N are inversely correlated across KCs, we asked if the number of PN inputs
315 per KC was inversely correlated to the number of synapses per PN-KC connection. We ignored PN-KC connections
316 with 2 or fewer synapses, because the number of synapses per PN-KC connection formed a bimodal distribution
317 with a trough around 3-4 (Fig. 7E); we presumed that connections with only 1-2 synapses represent annotation
318 errors. We divided KCs into their different subtypes as annotated in the hemibrain [45], because different subtypes
319 have different numbers of PN inputs per KC and different numbers of synapses per PN-KC connection ([28]; Fig.
320 7E,F, S5). We excluded KCs that receive significant non-olfactory input (γ -d, γ -t, $\alpha\beta$ -p, $\alpha'\beta'$ -ap1). In all analyzed
321 subtypes of KCs (γ -main, $\alpha\beta$ -s, -m and -c; $\alpha'\beta'$ -ap2 and -m), the number of PN inputs per KC (N) was inversely
322 correlated to the mean number of synapses per PN-KC connection, averaged across the PN inputs onto a KC (proxy
323 for \bar{w}) (Fig. 7G,K, S5). Linear regression showed that on average, there were $\approx 6 - 15\%$ fewer input synapses per
324 PN-KC connection (\bar{w}), for each additional PN per KC (N). This negative correlation meant that the number of
325 total PN-KC synapses per KC increased only sublinearly relative to the number of PN inputs per KC (Fig. S5).

326 We also tested another anatomical proxy of excitatory synaptic strength. Because KCs sum up synaptic inputs
327 linearly or sublinearly, their dendrites likely lack voltage-gated currents that would amplify inputs, so synaptic
328 input currents likely propagate passively [26]. Therefore, an excitatory input would make a smaller contribution to
329 a KC's decision to spike the farther away it is from the spike initiation zone [50]. While the spike initiation zone
330 cannot be directly observed in the connectome, the voltage-gated Na^+ channel *para* and other markers of the axon
331 initial segment (also called the 'distal axonal segment') are concentrated at the posterior end of the peduncle, near
332 where axons from KCs derived from the four neuroblast clones converge [51, 52]. This location can be approximated
333 in the connectome as the posterior boundary of the 'PED(R)' region of interest (ROI) (magenta dots, Fig. 7A,J).
334 From this point, we measured the distance along each KC's neurite skeleton (i.e., not the Euclidean distance) to
335 each PN-KC synapse. In the $\alpha\beta$ -c and γ -main KCs (but not other KCs), this distance was positively correlated
336 with the number of PNs per KC (Fig. 7H,K, S5). That is, the more PN inputs a KC has, the farther away the
337 input synapses are from the putative spike initiation zone (and thus the weaker they are likely to be). Intriguingly,
338 of all the KC subtypes, $\alpha\beta$ -c KCs show the strongest correlation between number of PN inputs and PN-peduncle
339 distance, but the weakest correlation between number of PN inputs and number of synapses per PN-KC connection
340 (Fig. 7K), suggesting that different types of KCs might use different mechanisms to achieve the same compensatory
341 end.

342 To test if inhibitory and excitatory input are positively correlated across KCs (as predicted in Fig. 7D), we
 343 approximated α by counting the number of synapses from the APL neuron to every KC in the calyx (the ‘CA(R)’
 344 ROI). In all types of KCs, the more total PN-KC synapses there were per KC, the more calyx APL-KC synapses
 345 there were (Fig. 7I,K, S5), indicating that indeed, inhibitory and excitatory synaptic input are correlated.
 346 These results confirm the predictions of our compensatory models. That correlations exist for both excitation
 347 and inhibition suggests that the mushroom body tunes more than one parameter simultaneously (thresholds may
 348 be tuned as well, but cannot be measured in the connectome). Such multi-parameter optimization likely explains
 349 (1) why the correlations in the connectome are not as steep as when only a single parameter is tuned in our models
 350 (Fig. 7D-F), and (2) why natural compensatory variation of tuned parameters need not be as wide as the variation
 351 of tuned parameters in our models (Fig. 5F).

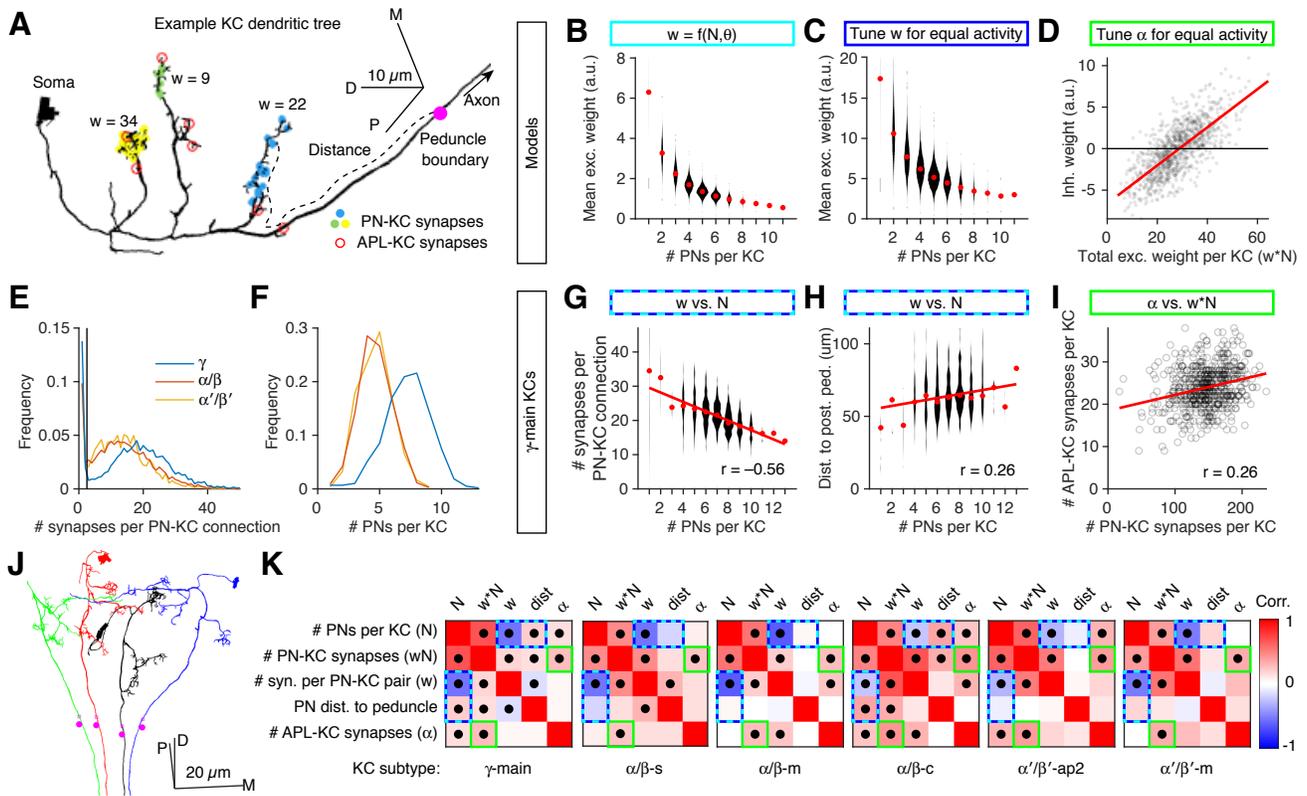


Figure 7: Connectome analysis reveals compensatory variation in excitatory and inhibitory input strengths. **(A)** Example $\alpha\beta$ -c KC (bodyId 5901207528) with inputs from 3 PNs (yellow/green/blue dots) and 7 dendritic APL-KC synapses (red circles). The magenta circle shows the posterior boundary of the peduncle. Line widths not to scale.

(B,C) Mean synaptic weight (w) per PN-KC connection is inversely related to the number of input PNs in models that tune input weights given N and θ **(B)**, or that tune input weights to equalize average activity levels across KCs **(C)**.

(D) In the model that tunes input inhibitory synaptic weights (α) to equalize average activity levels across KCs, inhibitory weights are directly related to the sum of excitatory weights per KC (i.e., wN). Note the negative values of α (discussed in text).

(E,F) Probability distributions of the number of synapses per PN-KC connection **(E)** and the number of input PNs per KC **(F)** in the different KCs subtypes ($\alpha\beta, \gamma, \alpha'\beta'$).

(G) Mean number of input synapses per PN-KC connection is inversely related to the number of input PNs per KC, in γ -main KCs (see Fig. S5 for other KC types).

(H) Mean distance of PN-KC synapses to the posterior boundary of the peduncle (presumed spike initiation zone) is directly related to the number of input PNs per KC.

(I) The number of APL-KC synapses per KC is directly related to the total number of PN-KC synapses per KC.

(J) Four $\alpha\beta$ -c KCs, one from each neuroblast clone. The posterior boundary of the peduncle (magenta circles) lies where the KC axons begin to converge.

(K) Grids show Pearson correlation coefficients (r) between various KC parameters for all KC subtypes tested (red: positive; blue: negative). Dots indicate $p < 0.05$ (Holm-Bonferroni corrected) (full statistics in Table S1). Colored outlines indicate predictions of models (cyan/blue: models tuning w **(G,H)**; green: model tuning α **(I)**). Number of KCs for each subtype, left to right: 588, 222, 350, 220, 127, 119. In (B,C,G,H), red dots are medians and the widths of the violin plots represent the number of KCs in each bin. Trend lines in (D,G,H,I) show linear fits to the data. Scale bars in **(A,J)**: D, dorsal, P, posterior, M, medial.

Discussion

Here we studied the computational costs and benefits of inter-neuronal variability for associative memory. Using a computational model of the fly mushroom body, we showed that associative memory performance is reduced by experimentally realistic variability among Kenyon cells in parameters that control neuronal excitability (spiking threshold and the number/strength of excitatory inputs). These deficits arise from the reduced separability and dimensionality of odor representations, which arises from unequal activity levels among Kenyon cells. However, memory performance can be rescued by using variability along one parameter to compensate for variability along other parameters, thereby equalizing average activity among KCs. These compensatory models predicted that certain KC features would be correlated with each other, and these predictions were borne out in the hemibrain connectome. In short, we showed (1) the computational benefits of compensatory variation, (2) multiple mechanisms by which such compensation can occur, and (3) anatomical evidence that such compensation does, in fact, occur.

Note that when we say "equalizing KC activity", we do not mean that all KCs should respond the same to a given odor. Rather, in each responding uniquely to different odors (due to their unique combinations of inputs from different PNs), they should keep their *average* activity levels the same. That is, while KCs' odor responses should be heterogeneous, their average activity should be homogeneous.

How robust are our connectome analyses? We found correlations between anatomical proxies for the physiological properties predicted to be correlated in our models (i.e., KCs receiving excitation from more PNs should have weaker excitatory inputs, while KCs receiving more overall excitation should also receive more inhibition). In particular, we measured the number of synapses per connection as a proxy for the strength of a connection. As described above,

373 this proxy seems valid based on matching anatomical and electrophysiological data [47–49]. However, other factors
374 affecting synaptic strength (receptor expression, post-translational modification of receptors, pre-synaptic vesicle
375 release, input resistance, etc.) would not be visible in the connectome. Of course, such factors could further enable
376 compensatory variability (see below), so anatomical proxies may actually underestimate the strength of correlations
377 between physiological properties.

378 We also used the distance between PN-KC synapses and the peduncle as a proxy for the passive decay of synaptic
379 currents as they travel to the spike initiation zone. In the absence of detailed compartmental models of KCs, it is
380 hard to predict exactly how much increased distance would reduce the effective strength of synaptic inputs, but it
381 is plausible to assume that signals decay monotonically with distance. Note that calcium signals are often entirely
382 restricted to one dendritic claw [26, 53]. Another caveat is that the posterior boundary of the peduncle is only an
383 estimate (though a plausible one: [51, 52]) of the location of the spike initiation zone. However, inaccurate locations
384 should only produce fictitious correlations for Fig. 7J and S5H if the error is correlated with the number of PN-KC
385 synapses per KC (and only in $\alpha\beta$ -c and γ -main KCs, not other KCs), which seems unlikely.

386 Our work is consistent with prior work, both theoretical and experimental, showing that compensatory variability
387 can maintain consistent network behavior [1–11, 54, 55]. However, to our knowledge, we are the first to analyze the
388 computational benefits of equalizing activity levels across neurons in a population (as opposed to across individual
389 animals or over time). A recent pre-print showed that equalizing response probabilities among KCs reduces memory
390 generalization [34], but we showed that equalizing average activity outperforms equalizing response probabilities
391 (Fig. S4), because only the former equalizes variance in activity among KCs to maximize dimensionality. Another
392 model of the mushroom body used compensatory inhibition, in which the strength of inhibition onto each KC
393 was proportional to its average excitation [31], similar to our inhibitory plasticity model (Fig. 5A2). However,
394 the previous work did not analyze the specific benefits from the compensatory variation; it also set the inhibition
395 strong enough that average net excitation was zero, whereas we show that when inhibition is constrained to be only
396 strong enough to reduce KC activity by \approx half (consistent with experimental data: [25]), inhibition alone cannot
397 realistically equalize KC activity (Fig. 5G). In addition, there is experimental support for our models' predictions
398 that KCs with more PN inputs would have weaker excitatory inputs: when predicting whether calcium influxes in
399 individual claws would add up to cause a supra-threshold response in the whole KC, the most accurate prediction
400 came from dividing the sum of claw responses by the log of the number of claws [53]. However, the functional
401 benefits of this result only become clear with our computational models. Finally, the larval mushroom body shows
402 a similar relationship between number and strength of PN-KC connections: the more PN inputs a KC has, the fewer
403 synapses per PN-KC connection [43]; however, again, the larval work did not analyze the computational benefits of
404 this correlation.

405 We modeled two forms of compensation: direct correlations between neuronal parameters (Fig. 5A1) and
406 activity-dependent homeostasis (Fig. 5A2-4). Both forms improve performance and predict observed correlations
407 in the connectome. We cannot directly resolve which mechanism explains the connectome correlations, but can

408 speculate by comparing whether key parameters are correlated with the number of PN inputs (N) but not total
409 number of PN-KC synapses ($\approx \bar{w}N$), which would suggest a mechanism based on dendritic morphology rather than
410 activity, or vice versa ($\bar{w}N$ but not N), which would suggest the opposite. Where PN-peduncle distance shows
411 significant correlations, it is correlated with both number of PN inputs and total number of PN-KC synapses, sug-
412 gesting that either mechanism is possible (Fig. 7). Conversely, the number of APL synapses ($\approx \alpha$) is more strongly
413 correlated with the total number of PN-KC synapses than with the number of PN inputs, which is more consistent
414 with activity-dependent tuning than parametric tuning. On the other hand, it may be that α is weakly directly
415 tuned to both \bar{w} and N and thus more strongly tuned to the combination, $\bar{w}N$.

416 Certainly, activity-dependent mechanisms are plausible, as KCs regulate their own activity homeostatically in
417 response to perturbations in activity [56]. Indeed, different KC subtypes use different combinations of mechanisms
418 for homeostatic plasticity [56], consistent with the different correlations observed in the connectome for different KC
419 subtypes. Our activity-dependent models lend themselves to straightforward biological interpretations. Excitatory
420 or inhibitory synaptic weights could be tuned by activity-dependent regulation of number of synapses per connection
421 or expression/localization of receptors or other post-synaptic machinery. Spiking thresholds could be tuned by
422 altering voltage-gated ion conductances or moving/resizing the spike initiation zone [52, 57].

423 On the other hand, KCs are not infinitely flexible in homeostatic regulation; for example, complete blockade
424 of inhibition causes the same increase in KC activity regardless of whether the blockade is acute (16 - 24 h) or
425 constitutive (throughout life) [56]. This apparent lack of activity-dependent down-regulation of excitation suggests
426 that activity-independent mechanisms might contribute to compensatory variation in KCs, as occurs for ion con-
427 ductances in lobster stomatogastric ganglion neurons [8, 9]. For example, the inverse correlation of w and N arises
428 from the fact that the number of PN-KC synapses per KC increases only sublinearly with increasing numbers of
429 claws (i.e., PN inputs) (Fig. S5H). Perhaps a metabolic or gene regulatory constraint prevents claws from recruiting
430 postsynaptic machinery in linear proportion to their number. (Interestingly, this suppression is stronger in larvae,
431 where the number of PN-KC synapses per KC is actually constant relative to the number of claws: [43].) Meanwhile,
432 the correlation between number of inhibitory synapses and number of excitatory synapses might be explained if
433 excitatory and inhibitory synapses share bottleneck synaptogenesis regulators on the post-synaptic side. Although
434 activity-dependent compensation produced superior performance in our model compared to activity-independent
435 compensation thanks to its more effective equalization of KC average activity (Fig. 5) (most likely because it takes
436 into account the unequal activity of different PNs), activity-dependent mechanisms suffered when the model net-
437 work switched to a novel odor environment (Fig. 6). Given that it is desirable for even a newly eclosed fly to learn
438 well, and for flies to learn to discriminate arbitrary novel odors, activity-independent mechanisms for compensatory
439 variation may be more effective in nature.

440 Compensatory variability to equalize activity across neurons could also occur in other systems. The vertebrate
441 cerebellum has an analogous architecture to the insect mushroom body; cerebellar granule cells are strikingly
442 similar to Kenyon cells in their circuit anatomy, proposed role in ‘expansion recoding’ for improved memory, and

443 even signaling pathways for synaptic plasticity [21, 40, 58–61]. Whereas cortical neurons’ average spontaneous
444 firing rates vary over several orders of magnitude [62], granule cells are, like Kenyon cells, mostly silent at rest, and
445 it is plausible that their average activity levels might be similar (while maintaining distinct responses to different
446 stimuli) [63]. Granule cell input synapses undergo homeostatic plasticity [64], while compartmental models suggest
447 that differences in granule cells’ dendritic morphology would affect their activity levels, an effect attenuated by
448 inhibition [65], raising the possibility that granule cells may also modulate inter-neuronal variability through activity-
449 dependent mechanisms. Future experiments may test whether compensatory variability occurs in, and improves the
450 function of, the cerebellum or other brain circuits. Finally, activity-dependent compensation may provide useful
451 techniques for machine learning. For example, we found that performance of a reservoir computing network could
452 be improved if thresholds of individual neurons are initialized to achieve a particular activity probability given the
453 distribution of input activities [66].

454 Methods

455 Modelling KC activity

PN activity was simulated using the odor responses of 24 olfactory receptors [30], passed through an equation proposed by [29]. For an ORN and PN innervating the i th glomerulus, their responses to the k th odor can be described using ORN_i^k (ORN activity) and x_i^k (PN activity):

$$x_i^k = R_{max} \frac{(ORN_i^k)^{1.5}}{(ORN_i^k)^{1.5} + (s^k)^{1.5} + \sigma^{1.5}} \quad (1)$$

where $s^k = m \sum_i ORN_i^k / 190$, $m = 10.63$, representing the gain of lateral inhibition in the antennal lobe, $R_{max} = 165$, representing the maximum PN response, and $\sigma = 12$, representing the non-linearity of the ORN-PN response function. We added noise to PN activity using:

$$(x_i^k)_{trial} = x_i^k (1 + CoVN) \quad (2)$$

456 where CoV is the coefficient of variation of PN activity across trials taken from Fig. 2E of [35] and \mathcal{N} is a random
457 sample drawn from a Gaussian distribution with mean 0 and standard deviation 1. To increase the number of
458 stimuli beyond the 110 recorded odors in [30], we generated odor responses in which the activity of each PN was
459 randomly sampled from that PN's activity across the 110 odors used in [30], i.e., $x_i^k = x_i^a$ where $k = 1 \dots K$, K being
460 the number of simulated odors, and a is randomly sampled from integers from 1 to 110 for each PN and each odor.

461 We modeled 2000 KCs. The j th KC received input from a randomly selected set of N_j PNs, where N_j was
462 either fixed at 6 or sampled from a Gaussian distribution with mean 6 and standard deviation 1.76 (integer values
463 only), based on experimental measurements from 200 KCs [28]. Although more recent results show that PN-
464 KC connectivity is not entirely random, as KCs that receive inputs from a certain group of food-odor-responsive
465 glomeruli are slightly more likely to receive other inputs from that same group [45, 67], we judged that attempting
466 to model this non-randomness would not add to the realism of our model given that we modeled only 24 (out of
467 ≈ 50) glomeruli.

The connection from the i th PN to the j th KC had strength w_{ji} , which was 0 for non-connected neurons, and for connected neurons was either fixed at 1, sampled from a log-normal distribution ($\mu = -0.0507$ and $\sigma = 0.3527$, based on [27]), or tuned by one of the methods described below. KCs received inhibition from APL (modeled as pseudo-feedforward for simplicity), with a gain that was either constant across all KCs (α) or tuned individually as described below (α_j). The KCs' spiking thresholds θ_j were either constant across all KCs, or sampled randomly from a Gaussian distribution with coefficient of variation 0.26, based on experimental measurements of the difference between spiking threshold and resting potential in 17 KCs [27]. These spiking thresholds were subject to a scaling factor C_θ to achieve the correct average coding level (see below). Thus, the activity of the j th KC for the k th odor,

y_j^k , was

$$y_j^k = \text{Relu}\left(\sum_{i=1}^{24} w_{ji}x_i^k - \alpha \sum_{j=1}^M \sum_{i=1}^{24} w_{ji}x_i^k - C_\theta\theta_j\right) \quad (3)$$

where $M = 2000$ is the number of KCs and Relu is a rectified linear unit:

$$\text{Relu}(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$$

The coding level, or fraction of KCs active for each odor, averaged across odors, was defined as:

$$CL = \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M H(y_j^k) \right] \quad (4)$$

where K and M are the number of odors and KCs, respectively and $H(x)$ is the Heaviside function:

$$H(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Experimental data suggest that coding level is around 0.1 normally, and approximately double that (0.2) when inhibition is blocked [25]. To match these constraints, we minimized this error function with respect to C_θ (thus preserving the coefficient of variation of thresholds across KCs, i.e., $C_\theta\theta_j$):

$$\epsilon_{CL|\alpha=0} = \frac{1}{2} [CL|_{\alpha=0} - CL_{target}|_{\alpha=0}]^2 \quad (5)$$

where $CL_{target}|_{\alpha=0} = 0.2$ and we minimized this error function with respect to α :

$$\epsilon_{CL} = \frac{1}{2} [CL - CL_{target}]^2 \quad (6)$$

468 where $CL_{target} = 0.1$.

We tuned C_θ and α using gradient optimization, using the update equations:

$$\Delta C_\theta = -\eta \frac{d\epsilon_{CL|\alpha=0}}{dC_\theta} \quad (7)$$

$$\Delta \alpha = -\eta \frac{d\epsilon_{CL}}{d\alpha} \quad (8)$$

To derive the update rule for ΔC_θ , we differentiate (5) with respect to C_θ :

$$\frac{d\epsilon_{CL|\alpha=0}}{dC_\theta} = [CL|_{\alpha=0} - CL_{target}|_{\alpha=0}] \frac{dCL|_{\alpha=0}}{dC_\theta} \quad (9)$$

To differentiate CL with respect to C_θ , we need to replace the discontinuous Heaviside function with a continuous approximation. Similar to [68] a sigmoid function approximates a Heaviside at the limit $\sigma \rightarrow 0$,

$$H(x) \approx S(x) = \frac{1}{1 + e^{-\frac{x}{\sigma}}} \quad (10)$$

Hence, assuming $\sigma = 1$, we can define the coding level as:

$$CL = \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M S(y_j^k) \right] \quad (11)$$

Given the derivative of a sigmoid is:

$$\begin{aligned} S'(x) &= \frac{dS(x)}{dx} = \frac{e^{-\frac{x}{\sigma}}}{[1 + e^{-\frac{x}{\sigma}}]^2} \\ &= S(x)(1 - S(x)) \end{aligned} \quad (12)$$

Thus,

$$\begin{aligned} \frac{dCL|_{\alpha=0}}{dC_\theta} &= \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M \left[S'(y_j^k |_{\alpha=0}) \frac{dy_j^k |_{\alpha=0}}{dC_\theta} \right] \right] \\ &= -\frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M [S'(y_j^k |_{\alpha=0}) H(y_j^k |_{\alpha=0}) \theta_j] \right] \end{aligned} \quad (13)$$

combining (9) and (13), and plugging in (7) we can get the update equation for C_θ as

$$\Delta C_\theta = \eta [CL|_{\alpha=0} - CL_{target}|_{\alpha=0}] \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M [S'(y_j^k |_{\alpha=0}) H(y_j^k |_{\alpha=0}) \theta_j] \right] \quad (14)$$

469 For simplicity, this can be re-written using the average operator notation $\langle \rangle$ across odors (indexed by k) and KCs
470 (indexed by j),

$$\Delta C_\theta = \eta [CL|_{\alpha=0} - CL_{target}|_{\alpha=0}] \langle S'(y_j^k |_{\alpha=0}) H(y_j^k |_{\alpha=0}) \theta_j \rangle_{j,k} \quad (15)$$

Similarly, for $\Delta\alpha$ we differentiate (6) with respect to α ,

$$\frac{d\epsilon_{CL}}{d\alpha} = [CL - CL_{target}] \frac{dCL}{d\alpha} \quad (16)$$

Similarly,

$$\begin{aligned} \frac{dCL}{d\alpha} &= \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M \left[S'(y_j^k) \frac{dy_j^k}{d\alpha} \right] \right] \\ &= -\frac{1}{K} \sum_{k=1}^K \left[\frac{1}{M} \sum_{j=1}^M \left[S'(y_j^k) H(y_j^k) \sum_j \sum_i w_{ji} x_i^k \right] \right] \end{aligned} \quad (17)$$

combining (16) with (17) then putting in (8),

$$\Delta\alpha = \eta [CL - CL_{target}] \frac{1}{MK} \sum_{k=1}^K \sum_{j=1}^M \left[S'(y_j^k) H(y_j^k) \sum_j \sum_i w_{ji} x_i^k \right] \quad (18)$$

and using the $\langle \rangle$ notation:

$$\Delta\alpha = \eta [CL - CL_{target}] \left\langle S'(y_j^k) H(y_j^k) \sum_j \sum_i w_{ji} x_i^k \right\rangle_{j,k} \quad (19)$$

471 These update equations were used to adjust values of θ and α in any random instantiation of the fly's network
 472 to match the experimentally observed coding levels. Note that because the update equation for α is the same for
 473 all j , the same equation applies when α_j is tuned for each KC (see below).

474 Modelling olfactory associative learning

Learning occurred through synaptic depression at the output synapse from KCs onto MBONs according to this exponential decay rule:

$$\Delta v_j = v_j (e^{-\eta y_j^k} - 1) \quad (20)$$

475 where v_j is the synaptic weight between the j th KC and the MBON of the 'wrong' valence and η is the learning
 476 rate. Thus, KCs active for a punished odor weaken their synapses to the approach MBON while KCs active for
 477 the rewarded odor weaken their synapses to the avoid MBON. This can be seen as the model fly learning from
 478 'mistakes' during its training phase [69, 70].

The behavior of the fly was determined by a softmax equation:

$$P(\text{approach}) = \frac{e^{cMBON_{\text{approach}}}}{e^{cMBON_{\text{avoid}}} + e^{cMBON_{\text{approach}}}} \quad (21)$$

479 where the constant c governs how probabilistic or deterministic the decision-making is. At high c , the model
 480 approaches a completely deterministic model where the fly will approach the odor 100% of the time whenever the
 481 approach MBON's activity is higher than the avoid MBON's activity; at very low c , the model approaches random
 482 chance; in between, the fly's behavior is probabilistic but biased by the imbalance between the activity of the two
 483 MBONs.

484 We trained the model on 15 noisy trials of the odors (no repetitions) and tested it on 15 unseen noisy trials of
 485 the same odors, and calculated the accuracy as the fraction of trials in which the model behaved correctly (i.e.,
 486 avoided punished odors and approached rewarded odors).

487 Metrics for evaluating Kenyon cell odor representations

Angular distance between two vectors A and B was calculated using:

$$\phi = \frac{2}{\pi} \arccos \frac{A \cdot B}{\|A\| \|B\|} \quad (22)$$

Dimensionality was calculated according to the equation in [40]:

$$dim(\mathbf{y}) = \frac{(\sum_{i=1}^m \lambda_i)^2}{\sum_{i=1}^m \lambda_i^2} \quad (23)$$

488 where λ_i are the eigenvalues of the covariance matrix of \mathbf{y} . Whereas Litwin-Kumar et al. calculated dimensionality
 489 analytically given inputs with defined distributions, we calculated it numerically given simulated PN inputs. Because
 490 dimensionality cannot be accurately calculated with a small number of inputs, we simulated KC activity for 1000
 491 input odors for dimensionality calculations.

Sparseness was calculated according to [25, 41]. Using the notation of this paper, the lifetime sparseness of the
 j th KC for a set of K odors is:

$$S_j = \frac{1}{1 - \frac{1}{K}} \left(1 - \frac{\left(\sum_{k=1}^K \frac{y_j^k}{K} \right)^2}{\sum_{k=1}^K \frac{(y_j^k)^2}{K}} \right) \quad (24)$$

492 If a cell is completely silent, firing to no stimuli, $y_j^k = 0$ for all k and sparseness is undefined due to division by zero.

We used the Davies-Bouldin Index (DBI; [39]) to measure the degree of separation between clusters of the KCs
 responses for two odors, or between the clusters of the rewarded odors responses versus the punished odors responses.
 The DBI measures the ratio between the within-cluster variance and the inter-cluster distance. Let clusters C_1 and
 C_2 consist of sets of A and B N -dimensional data points, $X = \{x_1, x_2, \dots, x_A\}$ and $Y = \{y_1, y_2, \dots, y_B\}$, respectively.
 The DBI is defined as:

$$DBI(C_1, C_2) = \frac{var(X) + var(Y)}{distance(\bar{X}, \bar{Y})} \quad (25)$$

where \bar{X} and \bar{Y} are the centroids of clusters C_1 and C_2 , and $distance(\bar{X}, \bar{Y})$ is the Euclidean distance between the
 two, while $var(X)$ and $var(Y)$ are the within-class variances, such that,

$$var(X) = \frac{1}{A} \sum_{i=1}^A (x_i - \bar{X})^2 \quad (26)$$

$$\text{var}(Y) = \frac{1}{B} \sum_{i=1}^B (y_i - \bar{Y})^2 \quad (27)$$

493 High DBI indicates poor separation (more overlap) between clusters C_1 and C_2 , due to either high within-cluster
494 variance or low inter-cluster distance.

495 Models for compensatory variability

496 Parametric tuning of excitatory input weights

497 We approximated the probability distribution of PN-KC synaptic weights (w) using the distribution of amplitudes
498 of spontaneous excitatory post-synaptic potentials (mini-EPSPs) in KCs, measured by [27]. This experimental
499 distribution was approximately log-normal, as has been described for cortical synapses [62, 71], so we modeled w as
500 following a log-normal distribution. We simulated values of w such that the overall distribution of w would follow
501 this log-normal distribution, yet individual KCs would sample w from different log-normal distributions depending
502 on N and θ , such that KCs with lower N or higher θ would have higher w , i.e., sampling from a log-normal
503 distribution shifted to the right (Fig. 5A1).

The probability of PN-to-KC synaptic weights could be estimated from the probability summation rule,

$$P(w) = \int_{\theta} \int_N P(w | N, \theta) P(N) P(\theta) dN d\theta \quad (28)$$

504 where $P(w | N, \theta)$ is the conditional probability distribution of the input synaptic weights for a KC that has N claws
505 and spiking threshold θ , sampled from probability distributions $P(N)$ and $P(\theta)$, respectively. We approximated
506 $P(N)$ and $P(\theta)$ as the Gaussian distributions described above (see Fig. 2), and we approximated integration over
507 θ as summation at small intervals ($\Delta\theta = 2.5$).

508 We modeled the constituent conditional probability distributions $P(w | N, \theta)$ as also being log-normal, based on
509 previous studies which approximate the sum of log-normal distributions as another log-normal variable by matching
510 the first two moments of the power sum and its individual log-normal contributors [72-74]. This approximation
511 holds in our case (the Kullback-Leibler Divergence metric (KLD) converged to less than 0.001).

To get the posterior lognormal distributions $P(w | N, \theta)$, we minimized the distance metric Kullback-Leibler
Divergence (KLD) between $P(w)$ and $\int_{\theta} \int_N P(w | N, \theta) P(N) P(\theta) dN d\theta$. To implement compensatory tuning in
these conditional probabilities, such that a KC with fewer inputs (lower N) or higher spiking threshold (higher θ)
would have stronger inputs (higher median w), we parameterized the medians $\tilde{\mu}$ of each conditional distribution in
 N and θ as:

$$\tilde{\mu} = \exp(\mu) = k \sqrt{\frac{\theta}{N}} \quad (29)$$

Thus,

$$\mu = \ln \left(k \sqrt{\frac{\theta}{N}} \right) \quad (30)$$

$$P(w | N, \theta) = \frac{1}{w\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\ln(w) - \ln\left(k\sqrt{\frac{\theta}{N}}\right)\right)^2}{2\sigma^2}\right) \quad (31)$$

We used gradient descent optimization to find the values of σ and k in Eq. 31 that would minimize the fitting error:

$$\begin{aligned} \epsilon &= KLD[P(w), \bar{P}(w)] \\ &= \int P(w) \ln \left[\frac{\bar{P}(w)}{P(w)} \right] dw \end{aligned} \quad (32)$$

where

$$\bar{P}(w) = \int_{\theta} \int_N P(w | N, \theta) P(N) P(\theta) dN d\theta \quad (33)$$

First, we found the optimal σ by gradient optimisation:

$$\Delta\sigma = -\eta_1 \frac{d\epsilon}{d\sigma} \quad (34)$$

The derivative of the fitting error with respect to σ is:

$$\frac{d\epsilon}{d\sigma} = - \int \frac{d\bar{P}(w)}{d\sigma} \frac{P(w)}{\bar{P}(w)} dw \quad (35)$$

with,

$$\frac{d\bar{P}(w)}{d\sigma} = \int_{\theta} \int_N \frac{dP(w | N, \theta)}{d\sigma} P(N) P(\theta) dN d\theta \quad (36)$$

where $\frac{dP(w|N,\theta)}{d\sigma}$ is:

$$\frac{dP(w | N, \theta)}{d\sigma} = \frac{1}{w\sigma^2\sqrt{2\pi}} \exp\left(-\frac{\left(\ln w - \ln\left(k\sqrt{\frac{\theta}{N}}\right)\right)^2}{2\sigma^2}\right) \left(\frac{1}{\sigma^2} \left(\ln w - \ln\left(k\sqrt{\frac{\theta}{N}}\right)\right)^2 - 1 \right) \quad (37)$$

Similarly for k ,

$$\begin{aligned} \Delta k &= -\eta_2 \frac{d\epsilon}{dk} \\ \frac{d\epsilon}{dk} &= - \int \frac{d\bar{P}(w)}{dk} \frac{P(w)}{\bar{P}(w)} dw \end{aligned} \quad (38)$$

such that,

$$\frac{d\bar{P}(w)}{dk} = \int_{\theta} \int_N \frac{dP(w | N, \theta)}{dk} P(N) P(\theta) dN d\theta \quad (39)$$

with $\frac{dP(w|N,\theta)}{dk}$ given by:

$$\frac{dP(w|N,\theta)}{dk} = \frac{1}{kw\sigma^3\sqrt{2\pi}} \exp - \frac{\left(\ln w - \ln\left(k\sqrt{\frac{\theta}{N}}\right)\right)^2}{2\sigma^2} \left(\ln w - \ln\left(k\sqrt{\frac{\theta}{N}}\right)\right) \quad (40)$$

512 Starting from arbitrary values for k and σ and using small learning rates η_1 and η_2 , at each iteration, the
 513 gradient descent algorithm alternated between using σ to update k and using k to update σ . We stopped the
 514 gradient descent (i.e., the algorithm converged) at $\epsilon < 0.001$.

515 **Tuning KC input excitatory weights to equalize KC activity**

In this model, we reduce the high variance in KCs' average activity levels by tuning their input synaptic weights, such that each j th KC adjusts its input synaptic weights (w_{ji}) to make its average activity level \bar{y}_j reach a certain desired level A_0 . We initially analyzed this problem using an error function:

$$\begin{aligned} \epsilon &= \frac{1}{2} [\bar{y}_j - A_0]^2 \\ \bar{y}_j &= \frac{1}{K} \sum_{k=1}^K y_j^k \end{aligned} \quad (41)$$

where y_j^k is the j th KC's response to the k th odor calculated as in equation (3) and K is the number of odors. Finding the weights to minimize the error in (41) can be found by gradient optimisation,

$$\Delta w_{ji} = -\eta \frac{d\epsilon}{dw_{ji}} \quad (42)$$

with,

$$\frac{d\epsilon}{dw_{ji}} = [\bar{y}_j - A_0] \frac{1}{K} \sum_{k=1}^K \frac{dy_j^k}{dw_{ji}} \quad (43)$$

Taking the derivative of y_j^k w.r.t. w_{ji} yields:

$$\frac{dy_j^k}{dw_{ji}} = H(y_j^k)(x_i^k - \alpha x_i^k) \quad (44)$$

Plugging (44) in (43) gives:

$$\begin{aligned} \frac{d\epsilon}{dw_{ji}} &= [\bar{y}_j - A_0] \frac{1}{K} \sum_{k=1}^K H(y_j^k)(x_i^k - \alpha x_i^k) \\ &= [\bar{y}_j - A_0] \langle H(y_j^k)(1 - \alpha)x_i^k \rangle_K \end{aligned} \quad (45)$$

Hence, w_{ji} will be updated as follows:

$$\Delta w_{ji} = -\eta [\bar{y}_j - A_0] \langle H(y_j^k)(1 - \alpha)x_i^k \rangle_K \quad (46)$$

The equation above means that a KC with an average activity \bar{y}_j higher (lower) than A_0 will scale down (up) its input synaptic weights, w_{ji} , proportional to both the difference $(y_j^k - A_0)$ and the average input activity from the i th PN. Note that in this derivation a KC must have non-zero average activity, i.e., $H(y_j^k) = 1$ for at least one odor, for its weights to be updated. We believe such a rule would be biologically implausible, as there should not be a discontinuity between a silent KC and a nearly silent KC. To allow totally silent KCs (which have only subthreshold activity) to update their weights in the same way as active KCs, we heuristically apply the following rule:

$$\Delta w_{ji} = -\eta [\bar{y}_j - A_0] \langle (1 - H(y_j^k))(1 - \alpha)x_i^k \rangle_K \quad (47)$$

Adding (46) and (47) we obtain:

$$\Delta w_{ji} = -\eta [\bar{y}_j - A_0] \langle (1 - \alpha)x_i^k \rangle_K \quad (48)$$

516 The rule has a fixed point $\bar{y}_j = A_0$ since $\langle (1 - \alpha)x_i^k \rangle_K > 0$. Note that we apply the constraint $w_{ji} \geq 0$. How
 517 updates for $w_{ji} = 0$ are treated depends on the reason why $w_{ji} = 0$: if the i th PN and j th KC are not connected,
 518 then the update is not applied. But if they were originally connected and the update rule pushed w_{ji} to zero, the
 519 update rule will continue to be applied.

520 To test whether performance is affected by adding the heuristic term to allow silent KCs to update their weights,
 521 we compared the performance using update rule Eq. (46) vs. (48). The rule without the heuristic performed
 522 significantly worse and had lower dimensionality than the rule with the added heuristic for activating silent KCs
 523 (Fig. S3A,B). This means that a formally derived update rule for w was not enough, since it would not equalize
 524 activity for all KCs (silent KCs will remain silent) and would not enhance the population coding as in the heuristic
 525 rule.

We further noted that Eq. (48) contains a factor x_i^k meaning that the update to w_{ji} depends on the average input activity from the i th PN. As this rule makes the biological interpretation more complex (the synaptic update depends on both pre- and post-synaptic activity), we also tested a simplified rule where synaptic changes depend only on the average KC activity:

$$\Delta w_{ji} = -\eta [\bar{y}_j - A_0] \quad (49)$$

526 This simplification did not affect memory performance or the tuned distribution of weights (Fig. S3A1,A2,D), but
 527 it improved the KCs' dimensionality (Fig. S3B) and the robustness of the model to novel odor environments (Fig.
 528 S3E). This improvement in the model robustness might be because including the extra factor x_i^k in the learning rule
 529 caused the model to be overfitted to the tuning environment. Therefore, we used Eq. (49) for the results presented

530 in the main figures, as it is simpler and produces better performance, despite not being formally derived from an
 531 error function. As with Eq. (48), this update rule has a fixed point $\bar{y}_j = A_0$.

532 **Tuning KC input inhibitory weights to equalize average KC activity**

In this model, we model each KC as adjusting its individual input inhibitory synaptic weights from APL, to match its average activity level \bar{y}_j to a certain desired level A_0 . We minimize the error function in Eq. (41) by adjusting α_j instead of w_{ji} :

$$\Delta\alpha_j = -\eta \frac{d\epsilon}{d\alpha_j} \quad (50)$$

$$\frac{d\epsilon}{d\alpha_j} = [\bar{y}_j - A_0] \frac{1}{K} \sum_{k=1}^K \frac{dy_j^k}{d\alpha_j} \quad (51)$$

Differentiating y_j^k with respect to α_j yields

$$\frac{dy_j^k}{d\alpha_j} = H(y_j^k) \left[-\sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right] \quad (52)$$

Plugging (63) in (51) gives,

$$\begin{aligned} \frac{d\epsilon}{d\alpha_j} &= [\bar{y}_j - A_0] \frac{1}{K} \sum_{k=1}^K H(y_j^k) \left[-\sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right] \\ &= [\bar{y}_j - A_0] \left\langle H(y_j^k) \left(-\sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right) \right\rangle_K \end{aligned} \quad (53)$$

Therefore,

$$\Delta\alpha_j = \eta [\bar{y}_j - A_0] \left\langle H(y_j^k) \left(\sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right) \right\rangle_K \quad (54)$$

Similar to the previous section, we assume that weight changes for silent neurons happen in the same way as for active neurons:

$$\Delta\alpha_j = \eta [\bar{y}_j - A_0] \left\langle (1 - H(y_j^k)) \left(\sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right) \right\rangle_K \quad (55)$$

533 .

Adding (54) and (55) we obtain the inhibitory plasticity rule allowing KCs to achieve equal average activity:

$$\Delta\alpha_j = \eta [\bar{y}_j - A_0] \left\langle \sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right\rangle_K \quad (56)$$

534 Given that $\left\langle \sum_j \sum_i w_{ji} x_i^k \right\rangle_K$ is a constant as w_{ji} is not updated in this model, this term can be subsumed into
 535 the learning rate, so this equation reduces to:

$$\Delta\alpha_j = \eta [\bar{y}_j - A_0] \quad (57)$$

536 Besides the homeostatic tuning of the APL inhibitory feedback values, these individual values of α_j also have
 537 to satisfy the sparsity constraint in Eq. (5). Therefore, the learning rule for these inhibitory weights requires
 538 simultaneously optimizing both error functions, Eq. (5) and (41). Thus combining Eq. (56) and the derivative of
 539 the sparsity constraint (CL=10%) with respect to each value of α_j ,

$$\Delta\alpha_j = \eta_1 [\bar{y}_j - A_0] - \eta_2 \frac{d\epsilon_{CL}}{d\alpha_j} \quad (58)$$

$$\Delta\alpha_j = \eta_1 [\bar{y}_j - A_0] - \eta_2 [CL - CL_{target}] \frac{dCL}{d\alpha_j} \quad (59)$$

540 where

$$\frac{dCL}{d\alpha_j} = -\frac{1}{MK} \sum_{k=1}^K \left[S'(y_j^k) H(y_j^k) \sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right] \quad (60)$$

Combining (59) with (60),

$$\Delta\alpha_j = \eta_1 [\bar{y}_j - A_0] + \eta_2 [CL - CL_{target}] \left\langle S'(y_j^k) H(y_j^k) \sum_{j=1}^M \sum_{i=1}^{24} w_{ji} x_i^k \right\rangle_k \quad (61)$$

541 We tested re-parameterizing α_j into $C_\alpha \alpha_j$ where C_α is tuned across all KCs to adjust coding level while α_j is
 542 tuned individually to equalize KC activity levels, but this had no effect on memory performance, so we kept the
 543 simpler model formulation.

544 Tuning KC spiking thresholds to equalize average KC activity

In this compensatory technique, we tune individual KCs' spiking thresholds θ_j to achieve equal average activity across the KC population. Starting with arbitrary initial values, each KC adjusts its spiking threshold so its average activity across K odors reaches a target level, A_0 , by minimizing the error in average activity as in Eq. (41) by gradient optimization:

$$\begin{aligned} \Delta\theta_j &= -\eta \frac{d\epsilon}{d\theta_j} \\ \frac{d\epsilon}{d\theta_j} &= [\bar{y}_j - A_0] \frac{1}{K} \sum_{k=1}^K \frac{dy_j^k}{d\theta_j} \end{aligned} \quad (62)$$

Differentiating y_j^k , the expression in Eq. (3), with respect to θ_j yields

$$\frac{dy_j^k}{d\theta_j} = H(y_j^k) [-C_\theta] \quad (63)$$

Plugging (63) in (62) gives,

$$\begin{aligned}\frac{d\epsilon}{d\theta_j} &= [\bar{y}_j - A_0] \frac{1}{K} \sum_{k=1}^K H(y_j^k) [-C_\theta] \\ &= -[\bar{y}_j - A_0] C_\theta \langle H(y_j^k) \rangle_k\end{aligned}\quad (64)$$

Therefore,

$$\Delta\theta_j = \eta [\bar{y}_j - A_0] C_\theta \langle H(y_j^k) \rangle_k \quad (65)$$

Similar to Eq. (47), we assume that spiking thresholds are updated for silent KCs as well:

$$\Delta\theta_j = \eta [\bar{y}_j - A_0] C_\theta \langle (1 - H(y_j^k)) \rangle_k \quad (66)$$

Adding (65) and (66) we obtain the spiking thresholds plasticity rule allowing KCs to achieve equal average activity:

$$\Delta\theta_j = \eta C_\theta [\bar{y}_j - A_0] \quad (67)$$

545 **Tuning spiking thresholds to equalize KCs response probabilities**

546 We tested an alternative strategy to tune θ suggested in [34]: to equalize not \bar{y}_j but rather the average response
547 probability of each KC across K odors without inhibition, P_j , i.e.:

$$P_j = \frac{1}{K} \sum_{k=1}^K H(y_j^k |_{\alpha=0}) \quad (68)$$

As in Eq. (5), we set this target response probability, $P_j^{target}|_{\alpha_j=0}$, to 0.2 to match experimental findings that blocking inhibition approximately doubles response probability [25]. We minimized the error function:

$$\epsilon = \frac{1}{2} [P_j - P_j^{target}|_{\alpha_j=0}]^2 \quad (69)$$

by adjusting θ_j by gradient optimization:

$$\begin{aligned}\Delta\theta_j &= -\eta \frac{d\epsilon}{d\theta_j} \\ \frac{d\epsilon}{d\theta_j} &= [P_j - P_j^{target}|_{\alpha_j=0}] \frac{dP_j}{d\theta_j}\end{aligned}\quad (70)$$

To differentiate P_j , as in Eq. (13), we approximated the discontinuous Heaviside function with a sigmoid:

$$\begin{aligned}\frac{dP_j}{d\theta_j} &= \frac{1}{K} \sum_{k=1}^K \frac{dS(y_j^k |_{\alpha=0})}{d\theta_j} \\ \frac{dS(y_j^k |_{\alpha=0})}{d\theta_j} &= S'(y_j^k |_{\alpha=0}) \frac{dy_j^k |_{\alpha=0}}{d\theta_j}\end{aligned}\quad (71)$$

Recalling the formula of y_j^k in (3), it follows

$$\frac{dy_j^k |_{\alpha=0}}{d\theta_j} = -C_\theta H(y_j^k) \quad (72)$$

Combining (72) with (71), and plugging in (70),

$$\frac{d\epsilon}{d\theta_j} = - [P_j - P_j^{target} |_{\alpha=0}] C_\theta \langle S'(y_j^k |_{\alpha=0}) H(y_j^k |_{\alpha=0}) \rangle_K \quad (73)$$

Thus, θ_j values are updated by,

$$\Delta\theta_j = \eta C_\theta [P_j - P_j^{target} |_{\alpha=0}] \langle S'(y_j^k |_{\alpha=0}) H(y_j^k |_{\alpha=0}) \rangle_K \quad (74)$$

As in Eq. (47), (66) and (55), we can write a symmetric rule for silent KCs:

$$\Delta\theta_j = \eta C_\theta [P_j - P_j^{target} |_{\alpha=0}] \langle S'(y_j^k |_{\alpha=0}) (1 - H(y_j^k |_{\alpha=0})) \rangle_K \quad (75)$$

Adding (75) and (74) leads to an activity-dependent update rule for θ_j , given all the incoming input odors:

$$\Delta\theta_j = \eta C_\theta [P_j - P_j^{target} |_{\alpha=0}] \langle S'(y_j^k |_{\alpha=0}) \rangle_K \quad (76)$$

In this model, the sparsity constraint $CL_{target|\alpha=0} = 0.2$ is satisfied by $P_j^{target} |_{\alpha_j=0} = 0.2$, because coding level equals the average of response probabilities across KCs:

$$\begin{aligned} CL &= \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{M} \sum_{j=1}^M H(y_j^k) \right) \\ &= \frac{1}{M} \sum_{j=1}^M \left(\frac{1}{K} \sum_{k=1}^K H(y_j^k) \right) \\ &= \langle P_j \rangle_j. \end{aligned} \quad (77)$$

548 Optimization of the multiple objective functions

549 As noted above, homeostatic tuning of w_{ji} , θ_j , or α_j needs to happen while maintaining the sparsity constraints,
550 Eq. (5) and (6). (It is important to note that the homeostatic update rules are meant to represent a biological
551 process while the sparsity constraints merely fit our model to experimental data and stand in for unknown processes
552 that lead to a coding level of 0.1.) Since these activity-equalizing tunings both depend on and change the network's
553 sparsity level, we used a sequential optimization approach to optimize each objective function, O_i , at a time. For
554 each i , we find the optimal parameters $\{P_i\}$ minimizing an objective O_i , using the current estimates of the other
555 parameters $\{P_j\}$ from all the other objectives, $\{O_j\}$ where $j \neq i$. The algorithm iterates for all i to minimise each
556 of the objective functions, until it reaches a global minimum where the errors from all of the objective functions

557 fall below a certain tolerance, τ_O .

558 Given an initial estimate for C_θ , α , θ_j and w_{ji} , the algorithm goes as follows:

Algorithm 1: Tuning of KCs parameters to equalize activity while constraining coding level

Result: C_θ , α , parameters to be tuned for activity equalization [w_{ji} or θ_j]

Initialize: [$C_\theta=1$, $\alpha=0$, $\epsilon_1=\epsilon_2=1$, $\epsilon_3=\bar{1}$, $\tau_1=0.2$ $\tau_2=0.01$, $\tau_3=0.06\mathbf{A}_0$]

Initialize tuned parameter for activity equalization [w_{ji} or θ_j] $\in U[0,1]$

while any in [$\epsilon_1, \epsilon_2, \epsilon_3$] > [τ_1, τ_2, τ_3] **do**

1. Using the current values for θ_j and w_{ji} , update C_θ using Eq.(15)

559 2. Using the value of C_θ from step (1) and current values for w_{ji} , and θ_j , update α using Eq. (19)

3. Using C_θ and α from (1) and (2) respectively, update w_{ji} using Eq. (46) or θ_j using Eq. (67)

4. Re-calculate the errors for the three objectives, Eq. (5), (6) and (41):

$$\epsilon_1 = \left| \frac{CL|_{\alpha=0}}{CL} - 2 \right|$$

$$\epsilon_2 = |CL - 0.1|$$

$$\epsilon_3 = |\bar{y}_j - \mathbf{A}_0|;$$

end while

560 In our implementation we initialize the parameters to be tuned for activity equalization (w_{ji} , θ_j or α_j) from a
 561 uniform random distribution $U = [0, 1]$ (the non-tuned parameters follow the distributions in Fig. 2). In addition,
 562 we set the error for the first and second sparsity constraint, Eq. (5) and (6), to be $\tau_1 = \left| \frac{CL|_{\alpha=0}}{CL} - 2 \right| = 0.2$, while $\tau_2 =$
 563 $|CL - 0.1| = 0.01$ respectively. This means allowing the coding level without and with the APL feedback to fall
 564 within $[1.8CL \leq CL|_{\alpha=0} \leq 2.2CL]$, and $[0.09 \leq CL \leq 0.11]$ respectively. For the activity equalization objective,
 565 the error ϵ_3 is a column vector of size M , of the differences between the target average activity value A_0 , and the
 566 current average activity for each KC, \bar{y}_j . This objective function is satisfied when all the values in the vector ϵ_3
 567 are less than 6% of the target activity.

568 Note that in the inhibition-tuning model, we tune the same parameter, α_j (a vector of M values instead of a
 569 constant), to jointly satisfy both the sparsity and the activity-equalization objectives. In this case, step (3) above
 570 is removed and step (2) updates α_j using Eq. (61).

571 In the model where we tune θ_j to equalize response probability rather than average activity (Fig. S4), equalizing
 572 response probability without inhibition to 0.2 also solves the coding level constraint (Eq. (77)). Thus, in this case,
 573 the algorithm iterates between 2 steps: (1) update θ_j according to Eq. (76), (2) use these values to update α
 574 according to Eq. (19), as follows,

Algorithm 2: Tuning of KCs spiking thresholds to equalize response probabilities

Result: C_θ , α , $[\theta_j]$ to be tuned for equalizing KCs response probabilities

Initialize: $[C_\theta=1, \alpha=0, \epsilon_1= \epsilon_2=1, \tau_1=0.2, \tau_2=0.01]$

Initialize $[\theta_j] \in U[0,1]$

while any in $[\epsilon_1, \epsilon_2] > [\tau_1, \tau_2]$ **do**

1. update θ_j using Eq. (76)
2. Using these new values of θ_j in step (1), update α using Eq. (19)
3. Re-calculate the errors for the two objectives, Eq. (69) and (6):

$$\epsilon_1 = |P_j - P_j^{target}|_{\alpha_j=0} |$$

$$\epsilon_2 = |CL-0.1 |$$

end while

In our optimization pipeline, there is a potential problem in the models where KC activity is equalized by tuning α_j or θ_j . In these models w_{ji} is not tuned, so for values of A_0 that are too high relative to values of w_{ji} , excitation will be too low to reach the high targets given the constraints $C_\theta\theta_j > 0$, $CL = 0.1$ and $CL|_{\alpha=0} = 0.2$, meaning the algorithm does not converge. (This is not a problem when tuning w_{ji} because w_{ji} can go arbitrarily high, whereas thresholds cannot go below zero.) Therefore, w_{ji} values must be chosen in a sensible range relative to A_0 (keeping in mind that the value of A_0 is arbitrary: see below). Rather than further complicating the objective cost functions by introducing a tunable scaling factor for w_{ji} , we found that in practice the algorithm converged if w_{ji} values (starting from a log-normal distribution with $\mu = -0.0507$, $\sigma = 0.3527$) were multiplied by $\frac{A_0}{CL}$ (where $CL = 0.1$). The target activity A_0 is arbitrary because if parameters can be found to satisfy our model constraints ($\bar{y}_j = A_0$, $CL = 0.1$ and $CL|_{\alpha=0} = 0.2$) for a particular $A_0 > 0$, then a solution also exists for $\bar{y}_j = cA_0$ for any $c > 0$, because:

$$\begin{aligned} cy_j^k &= c \operatorname{Relu}\left(\sum_{i=1}^{24} w_{ji}x_i^k - \alpha_j \sum_{j=1}^M \sum_{i=1}^{24} w_{ji}x_i^k - C_\theta\theta_j\right) \\ &= \operatorname{Relu}\left(\sum_{i=1}^{24} (cw_{ji})x_i^k - \alpha_j \sum_{j=1}^M \sum_{i=1}^{24} (cw_{ji})x_i^k - cC_\theta\theta_j\right) \end{aligned} \quad (78)$$

That is, to scale \bar{y}_j by a factor c , one need only scale the parameters w_{ji} and C_θ by c . In other words, only the relative magnitudes of A_0 , w_{ji} and C_θ , not the absolute magnitudes, are meaningful.

Robustness analysis

Of the 110 odors tested in [30], we took the four chemical classes with the most odors (acids, terpenes, alcohols and esters), so that tuning parameters on a single class would provide a reasonable number of odors (at least 15). Because each class had different numbers of odors, and the memory task is more difficult when more odors need to be classified, we equalized the number of odors in each task by randomly sampling 15 odors from those classes that had more than 15 members (terpenes, 16; alcohols, 18; esters, 24), with a different random sampling for each model instantiation. Because of the small number of odors used for tuning, it was not always possible to equalize

585 the activity of every single KC; in particular, in the threshold-tuning models in a novel environment, we allowed a
586 maximum of 5 KCs to fall outside the $\pm 6\%$ bound on average activity.

587 **Connectome analysis**

588 KC neurite skeletons and connectivity were downloaded from the hemibrain connectome v. 1.1 [44]. KCs (excluding
589 those that receive significant non-olfactory input) were selected as neurons whose ‘type’ field was *KCg-m*, *KCab-c*,
590 *KCab-m*, *KCab-s*, *KCa’b’-ap2* or *KCa’b’-m*. PN inputs for a KC were identified as neurons whose ‘type’ field included
591 *adPN*, *1PN* or *vPN* (NB: some of these, e.g., *vPNs*, do not project to the mushroom body and so were never counted)
592 and that formed more than 2 synapses with the KC (see Fig. 7B). KCs with truncated skeletons lacking the dendritic
593 tree were excluded. The posterior boundary of the peduncle was the most posterior node in a skeleton annotated as
594 being in the ‘PED(R)’ region of interest (annotations at [https://storage.cloud.google.com/hemibrain/v1.1/hemibrain-](https://storage.cloud.google.com/hemibrain/v1.1/hemibrain-v1.1-primary-roi-segmentation.tar.gz)
595 [v1.1-primary-roi-segmentation.tar.gz](https://storage.cloud.google.com/hemibrain/v1.1/hemibrain-v1.1-primary-roi-segmentation.tar.gz)). The boundary between the calyx and peduncle regions in the hemibrain was
596 defined by innervation by PNs (or lack thereof) (personal communication, K. Shinomiya). The distance from this
597 point to each PN-KC synapse along the KC’s neurite skeleton (i.e., not the Euclidean distance) was measured as
598 described in [36].

599 **Code availability**

600 Modeling and connectome analysis were carried out using custom code written in MATLAB, which is available at
601 <https://github.com/aclinlab/CompensatoryVariability>.

602 **Acknowledgements**

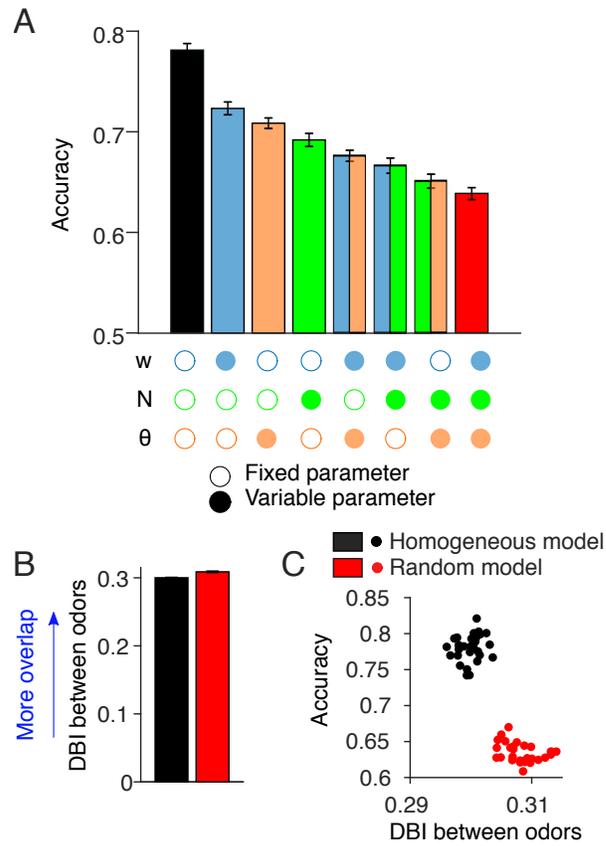
603 We thank Stuart Berg for ROI annotations in the connectome and Sofie Andersen, Caroni Fung, Celina Lubrino,
604 Charles McMillan, and Edirimuni Rodrigo for help in identifying truncated KC skeletons. This work was supported
605 by the European Research Council (639489 to ACL), the Biotechnology and Biological Sciences Research Council
606 (BB/S016031/1 to ACL), and the Engineering and Physical Sciences Research Council (2131691 to NA, EV and
607 ACL; EP/S009647/1, EP/S030964/1, EP/P006094/1 to EV).

608 **Author contributions**

609 NA, Conceptualization, Software, Formal analysis, Investigation, Visualization, Methodology, Writing - original
610 draft, Writing - review and editing. EV, Conceptualization, Formal analysis, Supervision, Funding acquisition,
611 Methodology, Writing - review and editing. ACL, Conceptualization, Software, Formal analysis, Supervision, Fund-
612 ing acquisition, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing.

613 **Supplementary Material**

614 S



615 **Figure S1:** Similar analyses to Fig. 2 and 3 with original odor responses from [30]. **(A)** Inter-KC variability degrades the memory performance when using the 110 odorants from [30]. **(B-C)** The Davies-Bouldin index is higher in the random model.

616

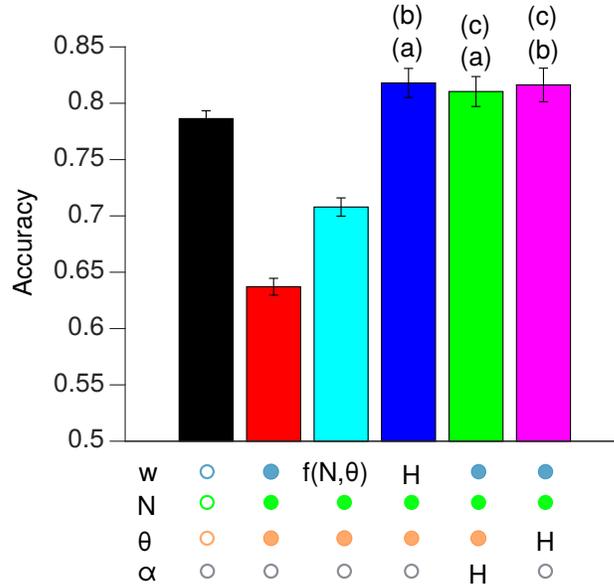


Figure S2: Similar analyses to Fig. 5 using the 110 odorants from [30]. The indeterminacy constant c from the softmax equation was set to 10. Bars that do not share the same letter annotation are significantly different, $p < 0.05$, Mann-Whitney or Wilcoxon test as in Fig. 5.

617

618

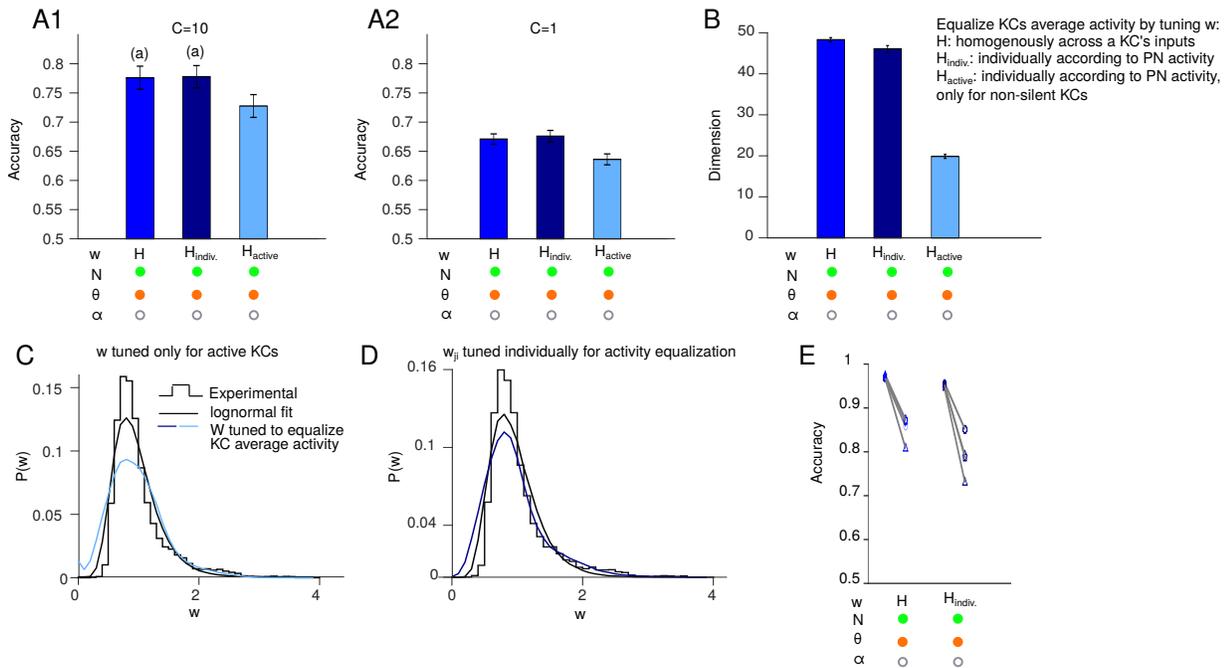


Figure S3: Alternative update rules for tuning KCs' input excitatory weights.

(A) Performance of different models at different indeterminacy constants (A1: $c = 10$; A2: $c = 1$): blue, the method in the main figures, Eq. (49), where a given KC's input weights are all adjusted equally ('H'); dark blue, Eq. (48), where a given KC's input weights are adjusted individually according to the average activity of the PN ('H_{indiv}'); light blue, Eq. (46), where only non-silent KCs adjust their input weights ('H_{active}').

(B) Dimensionality of KC odor representations. The 'H' model has a significantly higher dimensionality than both the 'H_{indiv}' and 'H_{active}' models. $n = 20$ model instances with different random PN-KC connectivity. Error bars show two times the SEM, i.e., 95.4% confidence interval. Bars with the same letter annotations are not significantly different from each other; all other comparisons are significant $p < 0.05$, by Wilcoxon signed-rank test with Holm-Bonferroni correction for multiple comparisons.

(C-D) Probability distribution of the tuned excitatory weights (compare to Fig. 5E).

619 (E) The 'H_{indiv}' model performs worse than the 'H' model in novel environments (see legend of Fig. 6).

620

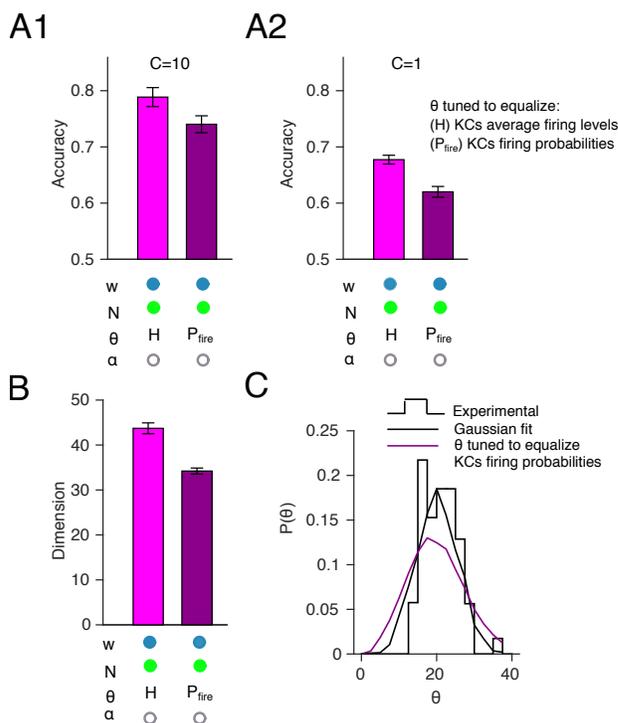


Figure S4: Equalizing KC average activity performs better than equalizing KC response probability.

(A) Better performance when spiking thresholds are tuned to equalize KC average activity (magenta) rather than KC response probability (dark magenta), under both more ($c = 10$, A1) and less ($c = 1$, A2) deterministic decision-making.

(B) Higher dimensionality of KC odor representations when equalizing KC average activity (magenta), compared to equalizing KC response probability (dark magenta). $n = 20$ model instances with different random PN-KC connectivity. Error bars show two times the SEM, i.e., 95.4% confidence interval. Magenta and dark magenta bars are significantly different, $p < 0.05$, by Wilcoxon signed-rank test.

621 (C) Probability distribution of spiking thresholds (θ) after tuning them to equalize KCs' response probabilities (compare to Fig. 5F).

622

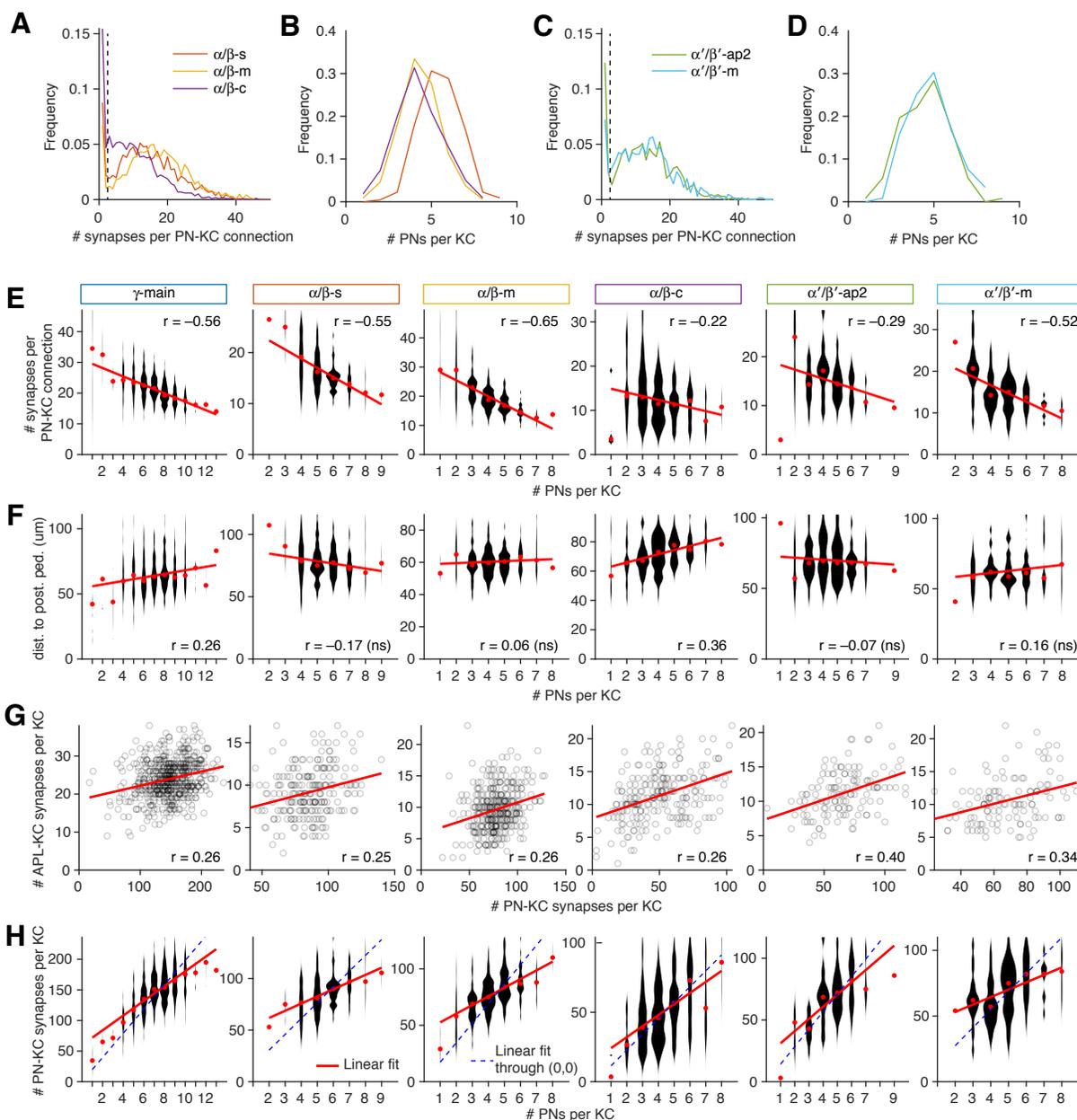


Figure S5: Connectome analysis on all KC subtypes (γ -main, $\alpha\beta$ -s, -m and -c; $\alpha'\beta'$ -ap2 and -m). **(A-D)** Probability distributions of the number of synapses per PN-KC connection (A,C) and the number of input PNs per KC (B,D) in $\alpha\beta$ and $\alpha'\beta'$ KCs separated out by subtype (compare to Fig. 7E,F). **(E)** Mean number of input synapses per PN-KC connection is inversely related to the number of input PNs per KC. **(F)** Mean distance of PN-KC synapses to the posterior boundary of the peduncle (presumed spike initiation zone) is directly related to the number of input PNs per KC in γ and $\alpha\beta$ -c KCs. **(G)** The number of APL-KC synapses per KC is directly related to the total number of PN-KC synapses per KC. **(H)** The number of PN-KC synapses per KCs grows sublinearly with the number of PN inputs per KC. Red dots: medians. Red lines: linear fits. Blue dashed lines: linear fits through the origin (if every PN-KC connection had the same number of synapses). Note that the red dots follow a concave function relative to both linear fits.

623

References

- [1] Jorge Golowasch et al. “Failure of Averaging in the Construction of a Conductance-Based Neuron Model”. In: *Journal of Neurophysiology* 87.2 (2002), pp. 1129–1131. DOI: <https://doi.org/10.1152/jn.00412.2001>.
- [2] Pablo Achard and Erik DeSchutter. “Complex Parameter Landscape for a Complex Neuron Model”. In: *PLOS Computational Biology* (2006). DOI: <https://doi.org/10.1371/journal.pcbi.0020094>.
- [3] Anne-Elise Tobin and Ronald L. Calabrese. “Endogenous and half-center bursting in morphologically-inspired models of leech heart interneurons”. In: *Journal of Neurophysiology* 96.4 (2006). DOI: 10.1152/jn.00025.2006.
- [4] Adam L. Taylor, Jean-Marc Goaillard, and Eve Marder. “How Multiple Conductances Determine Electrophysiological Properties in a Multicompartment Model”. In: *The Journal of Neuroscience* 29.17 (2009), pp. 5573–5586. DOI: 10.1523/JNEUROSCI.4438-08.2009.
- [5] Eve Marder and Jean-Marc Goaillard. “Variability, compensation and homeostasis in neuron and network function”. In: *Nature reviews Neuroscience* 7.7 (2006), pp. 563–74. DOI: 10.1038/nrn1949.
- [6] David J. Schulz, Jean-Marc Goaillard, and Eve Marder. “Variable channel expression in identified single and electrically coupled neurons in different animals”. In: *Nature Neuroscience* 9 (2006), pp. 356–362. DOI: <https://doi.org/10.1038/nn1639>.
- [7] David J. Schulz, Jean-Marc Goaillard, and Eve Marder. “Quantitative expression profiling of identified neurons reveals cell-specific constraints on highly variable levels of gene expression”. In: *PNAS* 32 (104 2007), pp. 13187–91. DOI: 10.1073/pnas.0705827104.
- [8] Jason N MacLean et al. “Activity-Independent Homeostasis in Rhythmically Active Neurons”. In: *Neuron* 37.1 (Jan. 2003), pp. 109–120.
- [9] Jason N MacLean et al. “Activity-independent coregulation of IA and Ih in rhythmically active neurons.” In: *Journal of Neurophysiology* 94.5 (Nov. 2005), pp. 3601–3617.
- [10] Timothy O’Leary and Eve Marder. “Temperature-Robust Neural Function from Activity-Dependent Ion Channel Regulation.” In: 26.21 (Nov. 2016), pp. 2935–2941.
- [11] Jay Z Parrish et al. “Krüppel Mediates the Selective Rebalancing of Ion Channel Expression”. In: *Neuron* 82.3 (May 2014), pp. 537–544.
- [12] Julijana Gjorgjieva, Guillaume Drion, and Eve Marder. “Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance”. In: *Current Opinion in Neurobiology* 37 (Apr. 2016), pp. 44–52.
- [13] Gary Marsat and Leonard Maler. “Neural Heterogeneity and Efficient Population Codes for Communication Signals”. In: *Journal of Neurophysiology* 104.5 (Nov. 2010), pp. 2543–2555.

- 656 [14] Fleur Zeldenrust, Boris Gutkin, and Sophie Denève. “Efficient and robust coding in heterogeneous recurrent
657 networks”. In: *bioRxiv* 10.6 (Oct. 2019), p. 804864.
- 658 [15] Krishnan Padmanabhan and Nathaniel N Urban. “Intrinsic biophysical diversity decorrelates neuronal firing
659 while increasing information content”. In: *Nature Neuroscience* 13.10 (Aug. 2010), pp. 1276–1282.
- 660 [16] Krishnan Padmanabhan and Nathaniel N Urban. “Disrupting information coding via block of 4-AP-sensitive
661 potassium channels”. In: *Journal of Neurophysiology* 112.5 (Sept. 2014), pp. 1054–1066.
- 662 [17] Shreejoy J Tripathy et al. “Intermediate intrinsic diversity enhances neural population coding”. In: *PNAS*
663 110.20 (May 2013), pp. 8248–8253.
- 664 [18] Luca Manneschi et al. *Exploiting Multiple Timescales in Hierarchical Echo State Networks*. 2021. arXiv:
665 2101.04223 [cs.LG].
- 666 [19] Nicolas Perez-Nieves et al. “Neural heterogeneity promotes robust learning”. In: *bioRxiv* 15.4 (Dec. 2020),
667 p. 2020.12.18.423468.
- 668 [20] David Oswald et al. “Activity of defined mushroom body output neurons underlies learned olfactory behavior
669 in *Drosophila*.” In: *Neuron* 86.2 (Apr. 2015), pp. 417–427.
- 670 [21] Annie Handler et al. “Distinct Dopamine Receptor Pathways Underlie the Temporal Sensitivity of Associative
671 Learning”. In: *Cell* 178.1 (2019), pp. 60–75. DOI: <https://doi.org/10.1016/j.cell.2019.05.040>.
- 672 [22] Toshihide Hige et al. “Heterosynaptic Plasticity Underlies Aversive Olfactory Learning in *Drosophila*”. In:
673 *Neuron* 88.5 (2015), pp. 985–998. DOI: 10.1016/j.neuron.2015.11.003.
- 674 [23] Hoger Amin and Andrew C. Lin. “Neuronal mechanisms underlying innate and learned olfactory processing
675 in *Drosophila*”. In: *Current Opinion in Insect Science* 36 (2019), pp. 9–17. DOI: [https://doi.org/10.1016/
676 j.cois.2019.06.003](https://doi.org/10.1016/j.cois.2019.06.003).
- 677 [24] Kyle S. Honegger, Robert A. A. Campbell, and Glenn C. Turner. “Cellular-Resolution Population Imaging
678 Reveals Robust Sparse Coding in the *Drosophila* Mushroom Body”. In: *Journal of Neuroscience* 31.33 (2011),
679 pp. 11772–11785. DOI: <https://doi.org/10.1523/JNEUROSCI.1099-11.2011>.
- 680 [25] Andrew C. Lin et al. “Sparse, Decorrelated Odor Coding in the Mushroom Body Enhances Learned Odor
681 Discrimination”. In: *Nature Neuroscience* 17.4 (2014), pp. 559–568. DOI: 10.1038/nn.3660.
- 682 [26] Eyal Gruntman and Glenn C. Turner. “Integration of the olfactory code across dendritic claws of single
683 mushroom body neurons”. In: *Nature Neuroscience* 16.12 (2013), pp. 1821–1829. DOI: 10.1038/nn.3547.
- 684 [27] Glenn C. Turner, Maxim Bazhenov, and Gilles Laurent. “Olfactory Representations by *Drosophila* Mushroom
685 Body Neurons”. In: *Journal of Neurophysiology* 99.2 (2008), pp. 734–746. DOI: [https://doi.org/10.1152/
686 jn.01283.2007](https://doi.org/10.1152/jn.01283.2007).
- 687 [28] Sophie Caron et al. “Random Convergence of Olfactory Inputs in the *Drosophila* Mushroom Body”. In: *Nature*
688 497.7447 (2013), pp. 113–117. DOI: 10.1038/nature12063.

- 689 [29] Shawn R. Olsen, Vikas Bhandawat, and Rachel I. Wilson. “Divisive normalization in olfactory population
690 codes”. In: *Neuron* 66.2 (2010), pp. 287–299. DOI: <https://doi.org/10.1016/j.neuron.2010.04.009>.
- 691 [30] Elissa A. Hallem and John R. Carlson. “Coding of Odors by a Receptor Repertoire”. In: *Cell* 125.1 (2006),
692 pp. 143–160. DOI: <https://doi.org/10.1016/j.cell.2006.01.050>.
- 693 [31] Sean X. Luo, Richard Axel, and L. F. Abbott. “Generating sparse and selective third-order responses in the
694 olfactory system of the fly”. In: *PNAS* 107.23 (2010), pp. 10713–10718. DOI: [10.1073/pnas.1005635107](https://doi.org/10.1073/pnas.1005635107).
- 695 [32] Moshe Parnas et al. “Odor Discrimination in *Drosophila*: From Neural Population Codes to Behavior”. In:
696 *Neuron* 79.5 (2013), pp. 932–44. DOI: [10.1016/j.neuron.2013.08.006](https://doi.org/10.1016/j.neuron.2013.08.006).
- 697 [33] K Krishnamurthy, A M Hermundstad, and T Mora. “Disorder and the neural representation of complex odors:
698 smelling in the real world”. In: *arXiv.org* (2017). scholar: C257FC5C-A5A2-4C43-8284-E9BA402F1DAA.
- 699 [34] Ann Kennedy. “Learning with naturalistic odor representations in a dynamic model of the *Drosophila* olfactory
700 system”. In: (2019). DOI: <https://doi.org/10.1101/783191>.
- 701 [35] Vikas Bhandawat et al. “Sensory Processing in the *Drosophila* Antennal Lobe Increases the Reliability and
702 Separability of Ensemble Odor Representations”. In: *Nature Neuroscience* 10.11 (2007), pp. 1474–1482. DOI:
703 [10.1038/nn1976](https://doi.org/10.1038/nn1976).
- 704 [36] Hoger Amin et al. “Localized inhibition in the *Drosophila* mushroom body”. In: *eLife* 9 (2020), e56954.
- 705 [37] Yoshinori Aso et al. “Mushroom body output neurons encode valence and guide memory-based action selection
706 in *Drosophila*”. In: *eLife* 3:e04580 (2014). DOI: <https://doi.org/10.7554/eLife.04580.001>.
- 707 [38] Raphael Cohn, Ianessa Morantte, and Vanessa Ruta. “Coordinated and Compartmentalized Neuromodulation
708 Shapes Sensory Processing in *Drosophila*”. In: *Cell* 163.7 (2015), pp. 1742–1755. DOI: <https://doi.org/10.1016/j.cell.2015.11.019>.
- 709
- 710 [39] David L. Davies and Donald W. Bouldin. “A Cluster Separation Measure”. In: *IEEE Transactions on Pattern
711 Analysis and Machine Intelligence-PAMI* 1 (2 1979), pp. 224–227. DOI: [10.1109/TPAMI.1979.4766909](https://doi.org/10.1109/TPAMI.1979.4766909).
- 712 [40] Ashok Litwin-Kumar et al. “Optimal Degrees of Synaptic Connectivity”. In: *Neuron* 93 (2017), pp. 1153–1164.
713 DOI: <https://doi.org/10.1016/j.neuron.2017.01.030>.
- 714 [41] B Willmore and D J Tolhurst. “Characterizing the sparseness of neural codes.” In: *Network: Computation in
715 Neural Systems* 12.3 (Aug. 2001), pp. 255–270.
- 716 [42] Kengo Inada, Yoshiko Tsuchimoto, and Hokto Kazama. “Origins of Cell-Type-Specific Olfactory Processing
717 in the *Drosophila* Mushroom Body Circuit.” In: *Neuron* 95.2 (July 2017), 357–367.e4.
- 718 [43] Katharina Eichler et al. “The complete connectome of a learning and memory centre in an insect brain.” In:
719 *Nature* 548.7666 (Aug. 2017), pp. 175–182.
- 720 [44] Louis K Scheffer et al. “A Connectome and Analysis of the Adult *Drosophila* Central Brain”. In: *bioRxiv* 12.4
721 (Apr. 2020), p. 2020.04.07.030213.

- 722 [45] Feng Li et al. “The connectome of the adult *Drosophila* mushroom body provides insights into function”. In:
723 *eLife* 9 (Dec. 2020).
- 724 [46] Quentin Gaudry et al. “Asymmetric neurotransmitter release enables rapid odour lateralization in
725 *Drosophila*.” In: *Nature* 493.7432 (Jan. 2013), pp. 424–428.
- 726 [47] William F. Tobin, Rachel I. Wilson, and Wei-Chung Allen Lee. “Wiring variations that enable and constrain
727 neural computation in a sensory microcircuit.” In: *eLife* 6 (May 2017), e24838.
- 728 [48] Christopher L Barnes, Daniel Bonnerly, and Albert Cardona. “Synaptic counts approximate synaptic contact
729 area in *Drosophila*”. In: *bioRxiv* (Oct. 2020), p. 2020.10.09.333187.
- 730 [49] Simone Holler et al. “Structure and function of a neocortical synapse”. In: *Nature* 314 (Jan. 2021), pp. 1–6.
- 731 [50] Stephen R Williams and Greg J Stuart. “Role of dendritic synapse location in the control of action potential
732 output”. In: *Trends in Neurosciences* 26.3 (Mar. 2003), pp. 147–154.
- 733 [51] Thomas A Ravenscroft et al. “*Drosophila* voltage-gated sodium channels are only expressed in active neurons
734 and are localized to distal axonal initial segment-like domains”. In: *Journal of Neuroscience* (Sept. 2020),
735 JN–RM–0142–20.
- 736 [52] Svetlana Trunova, Brian Baek, and Edward Giniger. “Cdk5 regulates the size of an axon initial segment-like
737 compartment in mushroom body neurons of the *Drosophila* central brain.” In: *The Journal of Neuroscience*
738 31.29 (July 2011), pp. 10451–10462.
- 739 [53] Hao Li et al. “Transformation of odor selectivity from projection neurons to single mushroom body neurons
740 mapped with dual-color calcium imaging.” In: *Proceedings of the National Academy of Sciences of the United*
741 *States of America* (July 2013).
- 742 [54] Rachel Grashow, Ted Brookings, and Eve Marder. “Compensation for variable intrinsic neuronal excitability
743 by circuit-synaptic interactions.” In: *The Journal of Neuroscience* 30.27 (July 2010), pp. 9145–9156.
- 744 [55] Hokto Kazama and Rachel I. Wilson. “Homeostatic matching and nonlinear amplification at identified central
745 synapses.” In: *Neuron* 58.3 (May 2008), pp. 401–413.
- 746 [56] Anthi A Apostolopoulou and Andrew C. Lin. “Mechanisms underlying homeostatic plasticity in the *Drosophila*
747 mushroom body in vivo.” In: *Proceedings of the National Academy of Sciences of the United States of America*
748 117.28 (July 2020), pp. 16606–16615.
- 749 [57] Matthew S Grubb and Juan Burrone. “Activity-dependent relocation of the axon initial segment fine-tunes
750 neuronal excitability.” In: *Nature* 465.7301 (June 2010), pp. 1070–1074.
- 751 [58] Mehrab N. Modi, Yichun Shuai, and Glenn C. Turner. “The *Drosophila* Mushroom Body: From Architecture
752 to Algorithm in a Learning Circuit.” In: *Annual Review of Neuroscience* 43.1 (July 2020), pp. 465–484.
- 753 [59] Sarah M Farris. “Are mushroom bodies cerebellum-like structures?” In: *Arthropod Structure and Development*
754 40.4 (July 2011), pp. 368–379.

- 755 [60] D Marr. “A theory of cerebellar cortex.” In: *The Journal of physiology* 202.2 (June 1969), pp. 437–470.
- 756 [61] Yoshinori Aso et al. “Nitric oxide acts as a cotransmitter in a subset of dopaminergic neurons to diversify
757 memory dynamics”. In: *eLife* 8 (Nov. 2019), e49257.
- 758 [62] György Buzsáki and Kenji Mizuseki. “The log-dynamic brain: how skewed distributions affect network oper-
759 ations”. In: *Nature Reviews Neuroscience* 15.4 (Apr. 2014), pp. 264–278.
- 760 [63] Kate Powell et al. “Synaptic representation of locomotion in single cerebellar granule cells”. In: *eLife* 4 (June
761 2015), p. 977.
- 762 [64] Igor Delvendahl, Katarzyna Kita, and Martin Müller. “Rapid and sustained homeostatic control of presynaptic
763 exocytosis at a central synapse.” In: *Proceedings of the National Academy of Sciences of the United States of*
764 *America* 116.47 (Nov. 2019), pp. 23783–23789.
- 765 [65] Catriona M Houston et al. “Exploring the significance of morphological diversity for cerebellar granule cell
766 excitability”. In: *Scientific Reports* 7.1 (Apr. 2017), pp. 1–16.
- 767 [66] Luca Manneschi, Andrew C. Lin, and Eleni Vasilaki. *SpaRCe: Improved Learning of Reservoir Computing*
768 *Systems through Sparse Representations*. 2021. arXiv: 1912.08124 [cs.NE].
- 769 [67] Zhihao Zheng et al. “Structured sampling of olfactory input by the fly mushroom body”. In: *bioRxiv* (Apr.
770 2020), p. 2020.04.17.047167.
- 771 [68] Xiaoying Han and Peter E. Kloeden. “Sigmoidal approximations of Heaviside functions in neural lattice
772 models”. In: *Journal of Differential Equations* 268.9 (2020), pp. 5283–5300. DOI: <https://doi.org/10.1016/j.jde.2019.11.010>.
- 773 [69] D R Chialvo and P Bak. “Learning From Mistakes”. In: *Neuroscience* 90.4 (1999), pp. 1137–48. DOI: [10.1016/s0306-4522\(98\)00472-2](https://doi.org/10.1016/s0306-4522(98)00472-2).
- 774 [70] James S Albus. “A theory of cerebellar function”. In: *Mathematical Biosciences* 10.1-2 (Feb. 1971), pp. 25–61.
- 775 [71] Sen Song et al. “Highly nonrandom features of synaptic connectivity in local cortical circuits.” In: *PLoS*
776 *biology* 3.3 (Mar. 2005), e68.
- 777 [72] L.F. Fenton. “The Sum of Log-Normal Probability Distributions In Scattered Transmission Systems”. In: *IRE*
778 *Trans.Commun.Systems* 8 (1960), pp. 57–67.
- 779 [73] S. Schwartz and Y.S. Yeh. “The Distribution Function and Moments Of Power Sums With Log-normal
780 Components”. In: *Bell Syst. Tech. J.* 61 (1982), pp. 1441–1462.
- 781 [74] Daniel Dufresne. “Sums of lognormals”. In: *Actuarial Research Conference*. 2008, pp. 1–6.