

Reading Out Olfactory Receptors: Feedforward Circuits Detect Odors in Mixtures without Demixing

Highlights

- A linear decoder can detect target odors in mixtures directly from glomerular input
- Odor detection can be achieved using only ~1% of the ~2,000 glomeruli
- A small training set (~100 trials) is sufficient to train the classifier
- Learning behavior of mice resembles discriminatively trained linear classifiers

Authors

Alexander Mathis, Dan Rokni,
Vikrant Kapoor, Matthias Bethge,
Venkatesh N. Murthy

Correspondence

vnmurthy@fas.harvard.edu

In Brief

Mathis et al. show that individual odors embedded in mixtures can be detected directly from olfactory receptor activity using simple linear classifiers with sparse weights. This result, and additional behavioral experiments, suggests that mice solve the mixture problem discriminatively.

Reading Out Olfactory Receptors: Feedforward Circuits Detect Odors in Mixtures without Demixing

Alexander Mathis,^{1,2,5} Dan Rokni,^{1,5} Vikrant Kapoor,¹ Matthias Bethge,^{2,3,4} and Venkatesh N. Murthy^{1,*}

¹Center for Brain Science and Department of Molecular & Cellular Biology, Harvard University, Cambridge, MA 02138 USA

²Werner Reichardt Centre for Integrative Neuroscience & Institute of Theoretical Physics, University of Tübingen, 72076 Tübingen, Germany

³Bernstein Center for Computational Neuroscience, University of Tübingen, 72076 Tübingen, Germany

⁴Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

⁵Co-first author

*Correspondence: vnmurthy@fas.harvard.edu

<http://dx.doi.org/10.1016/j.neuron.2016.08.007>

SUMMARY

The olfactory system, like other sensory systems, can detect specific stimuli of interest amidst complex, varying backgrounds. To gain insight into the neural mechanisms underlying this ability, we imaged responses of mouse olfactory bulb glomeruli to mixtures. We used this data to build a model of mixture responses that incorporated nonlinear interactions and trial-to-trial variability and explored potential decoding mechanisms that can mimic mouse performance when given glomerular responses as input. We find that a linear decoder with sparse weights could match mouse performance using just a small subset of the glomeruli (~15). However, when such a decoder is trained only with single odors, it generalizes poorly to mixture stimuli due to nonlinear mixture responses. We show that mice similarly fail to generalize, suggesting that they learn this segregation task discriminatively by adjusting task-specific decision boundaries without taking advantage of a demixed representation of odors.

INTRODUCTION

In natural environments, the olfactory system must be able to detect behaviorally relevant odors against varying background smells. This task resembles the cocktail party problem in audition where the signals of different sound sources also arrive as a linear mixture at the sensory organ. Similar to our remarkable ability of disentangling sound sources, mice have been shown to excel at the olfactory equivalent (Rokni et al., 2014). Yet, with an increasing number of mixture components this task can get difficult due to the overlapping, combinatorial representation of odorants by receptor neurons (Duchamp-Viret et al., 1999; Koulakov et al., 2007; Rospars et al., 2008; Shen et al., 2013) and sparse, distributed representations in higher olfactory

areas (Stettler and Axel, 2009; Wilson and Sullivan, 2011). These representations have been described as synthetic, in the sense that combinations of odorants are thought to be encoded rather than individual components (Gottfried, 2010; Wilson and Sullivan, 2011).

In the main olfactory system of the mouse, each sensory neuron expresses only one out of a family of ~1,000 types of olfactory receptor proteins (Godfrey et al., 2004). Each odor activates a subset of these receptor types and is identified by this subset, and each receptor type may be activated by many odors, leading to a potentially large overlap in the representation of different odorants (Duchamp-Viret et al., 1999; Malnic et al., 1999; Shen et al., 2013). Furthermore, mixtures may be represented as nonlinear functions of their parts (Rospars et al., 2008; Shen et al., 2013). In a previous study, we found that mice can be trained to perform a target-odor detection task and can do so remarkably well even in the presence of a large number of background odors (Rokni et al., 2014). These behavioral results led to the question of how mice solve this task given the nonlinear, noisy, and high-dimensional input that the glomeruli provide.

Various attempts have been made previously to provide theoretical understanding of how odors can be detected against a background. Several studies have suggested algorithms that rely on the temporal structure of OB activity to provide a concentration and background invariant code for odor identity (Brody and Hopfield, 2003; Galán et al., 2006; Hiratani and Fukai, 2015; Hopfield, 1991; Li, 1990). Detection of a specific component can also be achieved using Bayesian inference if one assumes prior knowledge of the receptor activation levels for “all odors.” Such prior knowledge could be implemented in the brain as a “generative model,” which could be formed by unsupervised learning (Mumford, 1994; Hinton and Sejnowski, 1999). Unsupervised learning could yield “demixed,” atomic representations of odors in higher olfactory areas—akin to Barlow’s idea that a lot of knowledge about the world is imprinted into the representations of sensory systems (Barlow, 1997). Based on a linear encoding model, Bayesian inference has been shown to work well in conditions where the typical mixture contains few components (Grabska-Barwinska et al., 2013). More generally, many algorithms for blind source separation and inference could be applied

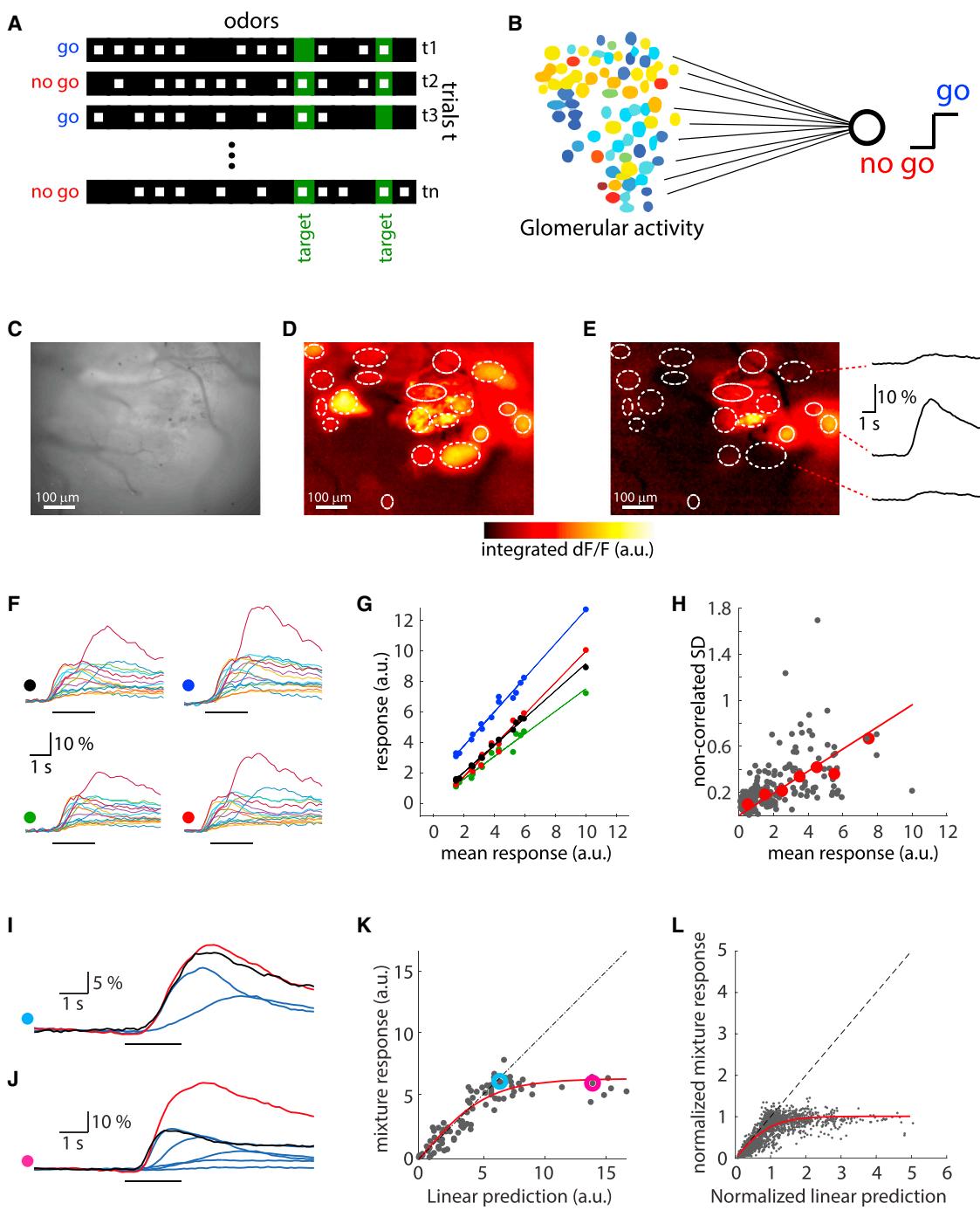


Figure 1. Variability and Summation of Glomerular Odor Responses

(A) Task structure. In each trial, an odor mixture is randomly generated from a pool of 16 odors. The task is to categorize based on the presence of one of two targets (green). Filled squares denote present odors, and empty squares denote absent odors.

(B) The classifier reads the glomerular responses as inputs and attempts to report the presence of the targets.

(C) Raw, temporal average fluorescence in one imaged area.

(D) Superposition of the dF/F responses to the 16 odors for same area as in (C). Ellipses show putative glomeruli.

(E) Average response to ethyl propionate. Time course of fluorescence dF/F changes shown for three glomeruli on the right.

(E) Average response to ethyl propionate. Time course of fluorescence GFP changes shown for (F) Responses to four presentations of methyl tiglate. Each panel shows one trial for all glomeruli.

(G) These four responses to methyl tiglate are plotted against the mean response. Colors represent trials as indicated by colored dots in (F). Lines are linear fits.

(G) These four responses to methyl tiglate are plotted against the mean response. Colors represent thals as indicated by colored dots in (F). Line (H) the trial-by-trial standard deviation after removal of correlated variability (see [Experimental Procedures](#)) is plotted against the mean response for each glomerulus-odor pair. Red dots are the median of mean-response-binned data. Red line is a linear fit to the data (forced through origin).

(legend continued on next page)

to such a generative model (Bell and Sejnowski, 1995; Otazu and Leibold, 2011; Tootoonian and Lengyel, 2014; Cuevas Rivera et al., 2015). Alternatively, supervised methods, like linear classifiers, could be used for discriminative learning (Galán et al., 2006; Berens et al., 2012; Shen et al., 2013). None of these studies, however, attempted to identify the decoding mechanism that mimics mouse psychophysical performance; this is the focus of the current work.

In this study, we sought to find the simplest decoding mechanism that uses experimentally constrained neuronal input and match the ability of mice to detect target odorants (Rokni et al., 2014). We measured representations of single odorants and hundreds of mixtures by olfactory receptors to derive a quantitative model of mixture responses, which also included saturation and experimentally measured trial-to-trial variability. We tested whether simple biologically plausible read-out mechanisms can solve the behavioral task based on inputs that were generated by the mixture model. We find that the simplest multi-glomerular approach, which sums up the weighted activity of different glomeruli and compares this sum to a threshold (a linear classifier), is sufficient to mimic the behavior of mice. Several glomeruli have to be combined, which argues against a strict labeled-line encoding model. The quality of the mixture samples used in training the model was important, since a model trained only on single odors only poorly generalizes to mixtures. We tested this prediction in mice and found a similar behavior.

RESULTS

In the behavioral task, mice were required to report the presence of one of two target odors in random mixtures containing up to 14 odors (Figure 1A). After learning, mice achieved accuracies close to 100% for a small number of distractors, and performance declined with increasing number of background odors (Rokni et al., 2014). This is a seemingly remarkable feat given that (a) mice had to generalize from around 1,000 training trials to mixtures that they had never smelled before (greater than 60% are novel in test phase; ~50,000 possible mixtures), (b) glomerular patterns are highly overlapping, (c) mixture responses arise from nonlinear interactions of single odors, and (d) odor responses are highly variable from trial to trial. To assess how challenging this task actually is, we started by quantifying all the parameters of the input at the level of receptors and then probed the ability of decoders to use glomerular activity patterns to extract the presence of particular target odors.

Estimating an Encoding Model for Mixtures Using Empirical Data

In earlier experiments (Rokni et al., 2014), we had measured the average glomerular responses to the 16 individual odors in

anesthetized mice expressing the Ca^{2+} indicator GCaMP3 in all olfactory receptor neurons (glomeruli were from the dorsal surface of the OB). The measured signals arise from the collection of sensory axons converging on glomeruli. The number of glomeruli that were responsive to at least one odor in each OB ranged between 63 and 76. Due to the large number of possible odor mixtures, measuring the neural responses to the whole stimulus set is intractable (>10,000 mixtures). Therefore, we estimated a statistical model that describes the glomerular responses given a particular odor mixture.

We imaged responses of glomeruli in the dorsal OB to all single components as well as a sample of ~500 mixtures in five additional OMP-GCaMP3 mice (Isogai et al., 2011; Figures 1C–1L). To estimate the variability of responses, each single odorant was presented several times per experiment (7.7 ± 1.7 , mean \pm SD; Figure 1F). Over all odor-glamorulus pairs, the average coefficient of variation (CV) was 0.37 ± 0.07 (mean \pm SD). However, much of the variability was correlated across glomeruli, probably reflecting causes that either are irrelevant to behaving mice (e.g., anesthesia level and breathing parameters; Blauvelt et al., 2013) or could be easily factored out by neural operations such as normalization. This correlated variability is evident when the responses of all glomeruli to each trial of a specific odor were plotted against the mean response to that odor across trials (Figure 1G). To estimate the non-correlated variability, we first subtracted the correlated variability from all responses. This was achieved by subtracting from each response pattern its best linear fit to the mean response pattern of the same odor (solid lines in Figure 1G; see Equation 4 in Experimental Procedures). The timescale of changes in the linear fit parameters was on the order of 10 min (Figure S1), suggesting factors such as variation in anesthesia. For each glomerulus i and odor j , we then calculated the uncorrelated component of the standard deviation (Equation 4) across trials t and plotted it against its mean response amplitude O_{ij} . The non-correlated CV for each experiment was calculated as the slope of the linear relationship between the non-correlated SD and response amplitude (Figure 1H). Non-correlated CVs ranged between 0.08 and 0.13 (0.099 ± 0.019 , mean \pm SD, $n = 5$ mice). This value is higher than expected based on the CVs measured in single rat receptor neurons *in vivo* (Duchamp-Viret et al., 2005), indicating that our measure of variability is probably conservative.

To estimate how glomerular responses to mixtures relate to their responses to mixture components, we imaged a sample of ~100 mixtures per experiment (98 ± 39 , mean \pm SD, $n = 5$) and compared the responses with the sum of the responses to individual components of the mixture (Figures 1I–1L). Most mixture responses could be estimated rather well by the linear sum of individual component responses (Figure 1I). However, beyond a certain value, mixture responses saturated (Figures

(I) An example of linear summation of mixture components. Shown are the average responses to individual components (blue), the response to the mixture (black), and the linear sum of the responses to the individual components (red).

(J) An example of sub-linear summation of mixture components. Colors are as in (I).

(K) All mixture responses are plotted against the linear sum of the responses to individual components for one single glomerulus. Colored data points show the examples in (I) and (J). Red line depicts the best sigmoid function fit (Equation 5; see Experimental Procedures).

(L) Same as (I) with data pooled across glomeruli and mice. Both the linear sum and the experimental data were normalized for each glomerulus to the saturation value of the fitted sigmoid function.

1J–1L). Saturation in odor responses may be due to several factors such as odor-receptor interaction, receptor neuron firing, and saturation of the Ca^{2+} indicator. We could not disambiguate which of these contribute to saturation in our dataset, and since receptor ligand interactions are well known to be saturating with increasing ligand concentration (Firestein et al., 1993), we conservatively assume that saturation is a real characteristic of odor responses of olfactory receptor neurons and use a saturating curve in our model for mixture responses.

These mixture responses can thus be summarized by the following encoding model that describes the activity vector R of glomerular responses in trial t (see [Experimental Procedures](#)):

$$R(t) = \lambda R_0(t) + (1 - \lambda)\sigma(R_0(t)). \quad (\text{Equation 1})$$

Here, $\lambda \in [0, 1]$ interpolates between the fully linear model $R_0(t)$ and the nonlinear model $\sigma(R_0(t))$, where σ denotes the saturating nonlinearity ([Equation 6](#)). The linear encoding part is given by

$$R_0(t) = \sum_{j=1}^{16} c_j(t) \cdot (O_j + \eta_j), \quad \eta_j \in \mathcal{N}(0, \alpha^2 O_j^2), \quad (\text{Equation 2})$$

where O_j is the glomerular response pattern for the j th odor, $c_j(t)$ is a binary variable denoting the presence of odor j in trial t , and the noise term η_j is normally distributed with a mean of zero and a variance of $\alpha^2 O_j^2$. The factor α controls the coefficient of variation, and our empirical estimate for it is ~ 0.1 . We will refer to the sequence of $c_j(t)$ during an experiment as the trial statistics.

In addition to introducing nonlinear mixing, parameterized by λ , we also systematically varied levels of trial-to-trial variability to study its effect on decoding performance.

Performance of Optimal Linear Decoder for Measured Mixture Encoding Model

The simplest decoder at the population level is a linear readout, similar to a single perceptron (Rosenblatt, 1958). Such a readout weighs the activity of individual glomeruli $R_i(t)$ by synaptic weights w_i and predicts the presence or absence of a target when the summed product is larger or smaller than a threshold θ , respectively. This linear–nonlinear cascade can be interpreted like a single readout neuron. Mathematically, the output is given by

$$y = H(w^T R(t) - \theta), \quad (\text{Equation 3})$$

where y is the binary output, H is the Heaviside function, which is 1 for positive arguments and 0 otherwise, and T denotes the transpose of the weight vector w .

In our published work (Rokni et al., 2014), there were 13 mice and 1,500–3,500 trials per mouse ($2,308 \pm 625$, mean \pm SD) with the number of odors per trial equally distributed between 1 and 14. For each mouse this segregation task can be summarized by the trial statistics $c_j(t)$ and the ideal behavior $r(t)$. For each animal's trial statistics, we calculated the optimal linear weights that minimize the mean squared deviations between the ideal behavior $r(t)$ and the output of the linear readout (see [Equation 8](#)) and the optimal decision threshold θ that maximizes performance on samples drawn from the encoding model ([Equation 1](#)),

for a random training set of 80% of the trials. This decoder, which we call the OLE (optimal linear estimator), was then tested on the remaining 20% of trials ([Experimental Procedures](#)). We repeated this procedure 20 times and report the average performance on the test sets.

Performance was perfect without saturation ($\lambda = 1$) and trial-to-trial variability ($\alpha = 0$) ([Figure 2B](#)). Increasing α gradually decreased performance, which reached $\sim 65\%$ for $\alpha = 1$ ([Figures 2A and 2B](#)). Similar trends were obtained using the glomerular patterns from different imaging datasets ([Figure 2A](#)) and target odors ([Figure 2B](#)). For a noise level similar to our experimental estimate (~ 0.1), most linear readouts performed better than corresponding mice detecting the same targets ([Figure 2B](#)). Thus, considering only trial-to-trial variability and mixing of odors without saturation, a simple linear decoder is sufficient to match or exceed mouse performance, although it had access only to the time-averaged response of a limited set of dorsal glomeruli. Since the specific choice of data had little effect on classifier performance, we focused on a particular imaging set with 72 glomeruli.

Next, we varied glomerular saturation by varying the parameter λ in [Equation 1](#). As expected, the average performance of the optimal linear decoder for all target odors declined with both increasing nonlinearity and noise ([Figure 2C](#), single odor pair [Figure S2A](#)). When noise levels are low (< 0.1) and responses are not fully saturating ($\lambda > 0$), the mean decoder performance is above 98%. When considering saturating glomeruli ($\lambda = 0$) and the experimental estimate of noise ($\alpha = 0.1$), performance drops to 91.6%, which is comparable to the average mouse performance of 90.3%. This demonstrates that the “cocktail party problem” is almost linearly separable at the level of glomeruli.

Performance of Support Vector Machine for Noisy, Nonlinear Mixtures

Since the glomerular activity patterns in the linear, noisy mixture case $R(t)$ are given by a superposition of multiple Gaussian random vectors, the optimal decision boundary might be highly nonlinear and not merely a single hyperplane. When the odor patterns are nonlinearly mixed, the geometry of the decision boundaries becomes even more intricate. A decoder that is able to flexibly adjust its decision boundaries without constraining them to be a single hyperplane may therefore perform considerably better than the simple linear readout. One powerful class of decoders with such properties is the SVM, which uses a so-called kernel trick. For SVMs that use radial basis functions as kernels, the decision boundaries are formed by sums of Gaussians and can, in the limit, approximate almost any decision boundary (Schoelkopf and Smola, 2002).

We trained SVM decoders with the same cross-validation procedure as described for the optimal linear decoder. As expected from their versatility, SVMs were more accurate than the linear readout. [Figure 2D](#) shows the performance of the SVM for the same target odors as the optimal linear decoder in [Figure 2C](#). When noise levels are low ($\alpha < 0.1$) and responses are not fully saturating ($\lambda > 0$), the SVM performance was above 99.5%. When considering saturating glomeruli ($\lambda = 0$) and the experimental estimate of noise ($\alpha = 0.1$), performance drops to 94.7%, outperforming mice and making about 40% fewer

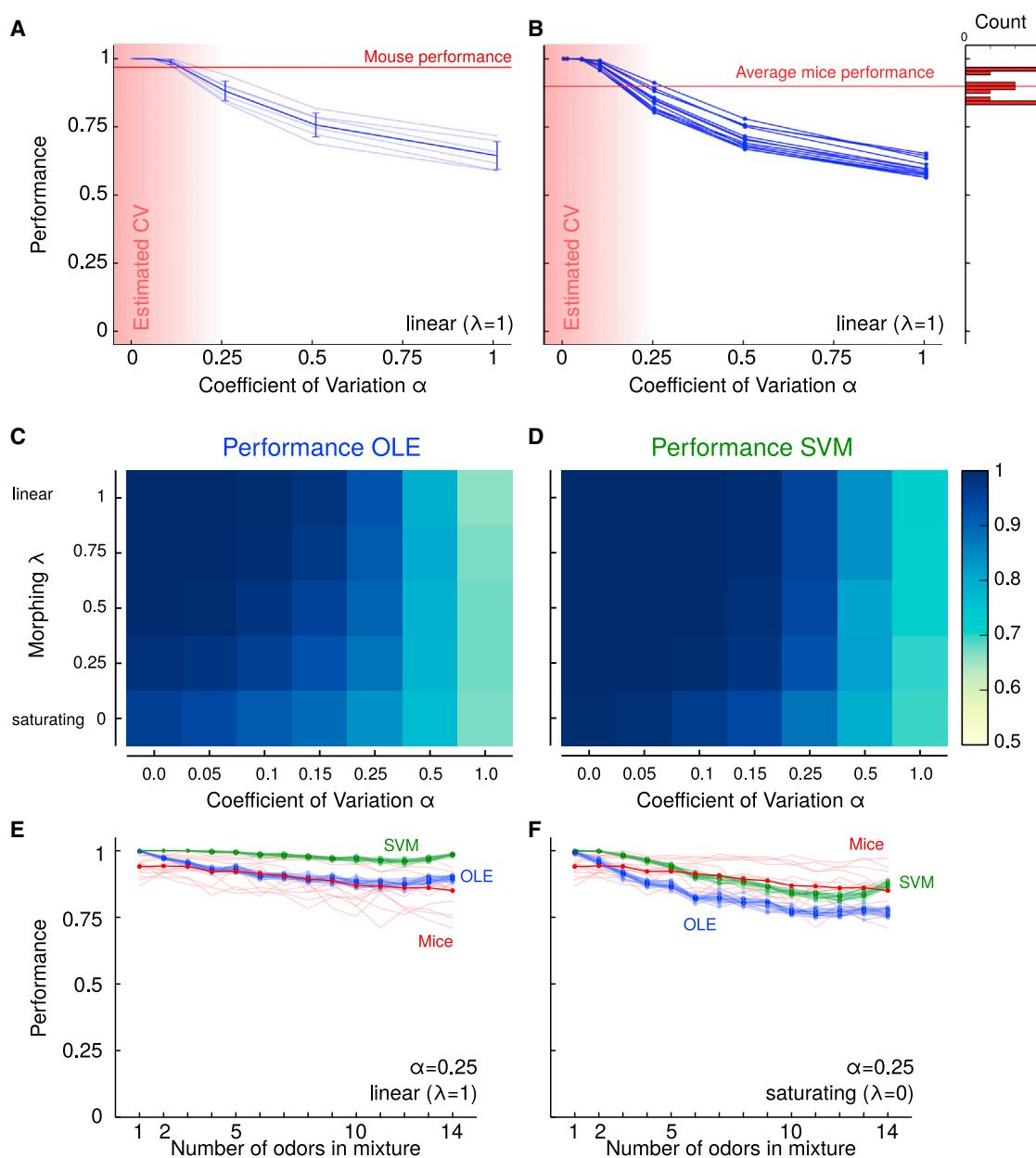


Figure 2. Decoding Performance of Linear Readout and SVM for Mixtures with Trial-to-Trial Variability and Nonlinear Mixtures

- (A) Performance of optimal linear readout in detecting one target pair (isobutyl propionate and allyl butyrate) is plotted as a function of trial-to-trial variability (α). The curves show the performance of one decoder based on each of the six measured glomerular maps and their average \pm SD. The mouse performance for these targets is shown as a horizontal line. The estimated range of the glomerular noise is highlighted by the shaded red region.
- (B) Average decoding performance for all 13 pairs of target odors. The average performance of all mice is indicated by a horizontal line, and the distribution of experimental performances is shown to the right.
- (C) Average performance for data from all mice of optimal linear decoder for varying noise levels and nonlinearities from $\lambda = 1$ (linear case) to $\lambda = 0$ (experimentally measured saturation).
- (D) Average performance for data from all mice, using SVM with radial basis function for mixture data as in (C).
- (E) Performance of OLE and SVM as a function of the number of odors in the mixture. Performance curves for 13 mice and their average in red. The average performance curves for the SVMs and OLEs are shown for the same 13 target odor pairs for linear mixture model ($\lambda = 1$) with $\sigma = 0.25$. Those are the conditions for which the experimental and model performance curves are best correlated (Figures S2C and S2D).
- (F) Same as (E), but for mixtures with saturation ($\lambda = 0$).

mistakes than the optimal linear decoder. Similarly, for other conditions, SVMs make fewer errors than the optimal linear decoder and have substantially better performance for low noise levels irrespective of λ (Figures 2C and 2D).

Given that the overall performance of both optimal linear decoder and SVM broadly resembled that of mice, we investigated whether they make similar errors. Specifically, we asked whether the dependence of performance on the number of mixture components is similar to the monotonic decline seen in mice (Figures 2E and 2F). We refer to this relationship as the performance curve. We calculated the average correlation coefficient (r) between the performance curves of the classifiers for specific odor targets and their corresponding mice. The average correlation was maximal for the fully saturated mixture model ($\lambda = 0$) with a noise level of $\alpha = 0.25$ for both the optimal linear decoder ($r = 0.42 \pm 0.01$, mean \pm SEM, $n = 13$ mice) and SVM ($r = 0.40 \pm 0.01$, mean \pm SEM, $n = 13$ mice; Figures S2C and S2D). When scrutinizing the shape of the performance curves, it is important to note that the number of possible mixtures varies with the number of components in the mixture. For instance, there are many more possible mixtures of 7 components than there are of 1 or 14 components. Therefore, specific stimuli on the edges of the performance curve repeat much more, offering more training samples for the classifier. For this reason, one may expect that classifier (and mouse) performance curves will not be monotonic, but could have a minimum at the midrange of mixtures. We find, however, that despite the different computational capabilities of SVMs and the OLE, performance for both decoders decreases as the number of components in a mixture increase, for realistic estimates of noise and nonlinearity. This suggests that the overlapping and nonlinear mixing of noisy glomerular patterns, rather than limitations of the readout mechanism, is a fundamental bottleneck for this task.

Taken together, although the linear decoder is not the optimal decoder, its performance is comparable to that of mice for our conservative estimates of noise and saturation levels. We therefore used this linear decoder, rather than the SVM, for further analysis. Nevertheless, it is conceivable that nonlinear decoding schemes such as the SVM could be good models for certain olfactory computations, but for this task, nonlinear decision boundaries are not necessary.

Performance of Sparse, Linear Decoder for Noisy, Nonlinear Mixtures

The linear decoders considered so far had no restrictions on the number of glomeruli used. Inspired by the observation that the projections from the OB to the piriform cortex are sparse and disperse (Giessel and Datta, 2014), we wondered how well sparse, linear readouts would perform the task. We turned to logistic regression, which is an efficient, binary classification algorithm that can be readily interpreted from a neuronal point of view as a linear readout neuron with nonlinear firing (Berens et al., 2012). Similar to the OLE, the optimal logistic regression (OLR) is found by minimizing the likelihood of misclassifications on a training set. Additionally, we employed a regularization term that punishes non-zero readout weights by adding the sum of all absolute values of the weights to the number of

misclassifications (L1 – regularization; Equation 10). By weighing this regularization term with a constant $1/C$ that can be systematically varied, one can bias the OLR to exhibit varying degrees of sparseness. This regularization can be thought of as an additional cost term for neuronal wiring, and the OLR attempts to maximize performance while minimizing wiring.

When the cost for wiring dominates, the performance of the OLR drops to chance level and at the other extreme performance asymptotes to levels similar to the optimal linear estimator. Between these two extremes, performance strongly depends on the number of nonzero readout weights (Figure 3A). However, the overall effect of readout sparseness on performance was mild; this is not a simple consequence of the glomerular activity itself being sparse, since most glomeruli are activated by at least one odor. For our estimated noise level of 0.1, only 10–20 glomeruli are sufficient to match the performance of mice. To reach 90% of the asymptotic performance, only 15.7 ± 0.9 glomeruli (mean \pm SEM, $n = 13$ mice) were necessary (Figure 3A). The OLR classifier is trained merely based on the sampled mixtures vectors $R(t)$, and the ideal response for the fraction of training trials, but has no explicit knowledge of the identity of the target odor or the glomerular representations of the individual components. Due to the large number of possible mixtures, even mice that were trained on the same targets were exposed to different trial statistics. Despite these differences, classifiers trained on trial statistics from different mice with the same target odors converged to similar weights in both the sparse and the dense case (Figure 3B, see odor pairs F, G, and H). When mice had different targets, the readout weights were also distinct (Figure 3B). These weights are not simply a reflection of the glomerular patterns of the target odors, but they also take into account glomerular patterns of distractors (Figure S3A). Using the target-odor patterns as readout weights results in poor performance; even for just one target odor, such a template matching algorithm is about 20% worse (Figures S3B and S3C).

Saturation also affects which and how many glomeruli are read out by the OLR (Figures S4A and S4B). For linear mixing ($\lambda = 1$), only ~ 10 glomeruli are necessary to achieve 90% performance (10.6 ± 0.5 , mean \pm SEM, $n = 13$ mice). These glomeruli are not just a subset of the ones used in the fully saturated case.

Multiple glomeruli are necessary for detecting the target odors from noisy and saturating glomerular inputs. We asked whether single glomeruli may be sufficient to detect the targets making less conservative assumptions. We used receiver operating characteristic (ROC) analysis, a standard technique in signal detection, to estimate the ability to decode using single glomeruli, assuming linear summation of glomerular inputs and no trial-to-trial variability. We found that even in these favorable conditions, the best glomerulus-target odor pairing leads to a performance of less than 75% (with top first percentile 68.0%, and mean \pm SD performance $52.8\% \pm 0.04\%$, $n = 5,473$ glomerulus-odor targets pairs; Figures S4C and S4D).

Since synaptic transmission is inherently noisy in the nervous system (Del Castillo and Katz, 1954), an important question is how robust the decoder's performance is to fluctuations of the readout weights. To test for robustness, we perturbed the classifier weights (for a dense readout, $C = 10^6$) by multiplying them by a random factor that is normally distributed with a mean of

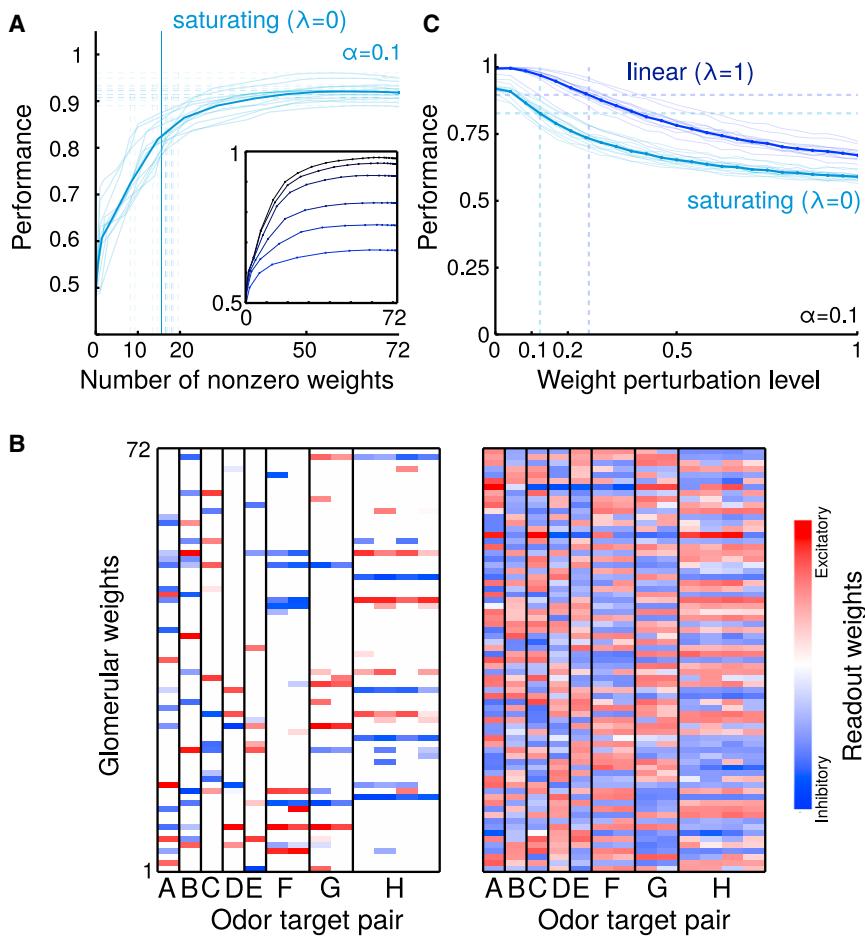


Figure 3. Performance of Sparse Linear Readout

(A) Performance versus number of non-zero weights of the OLR ($\lambda = 0$). Each faded line is for data from one animal, and solid line depicts the average. The horizontal, dashed lines indicate asymptotic performance, and the vertical dashed lines indicate the corresponding minimal number of glomeruli to achieve 90% of that performance per animal. Their average is highlighted as a solid vertical line. Inset: average performance versus number of non-zero weights of the OLR for varying noise levels ($\alpha = 0, 0.05, 0.1, 0.25, 0.5, 1$, from top to bottom).

(B) Example readout weights obtained for 90% of asymptotic performance (left) and asymptotic performance (right). Several of these 13 mice had the same target odor pairs, as indicated by grouping under the same letter. OLR learns similar readout weights from different trial statistics for the same target odors. Across different target odors, the readout weights vary substantially, but the weights are similar for the same target odors. (C) Performance of OLR versus perturbation level of weights for linear and nonlinear mixing with $\alpha = 0.1$. Each thin line is the average for 20 random perturbations per mouse, and the thick line is the average across mice. The horizontal lines indicate 90% of the unperturbed weight performance for either condition ($\lambda = 0, 1$). The vertical lines indicate the perturbation level where the performance drops below 90% of the unperturbed weight performance.

one. We varied the perturbation level by varying the standard deviation of this factor. We found that the linear decoder is fairly resilient to changes in its readout weights and that robustness is higher for the linear mixing case ($\lambda = 1$; Figure 3C). Robustness of classifier performance strongly depends on the specific target odors, yet it remained above 90% of optimal for all 13 trial statistics at perturbation levels that are less than 0.25 for $\lambda = 1$ and 0.1 for $\lambda = 0$.

Learnability and Experience Dependency of Readout Weights

We established that this task is linearly separable, but even a linear decoder has at least as many degrees of freedom as the number of mouse glomeruli. This raises the question of how difficult it is to learn the readout weights; that is, how many examples are needed to properly constrain the weights? To quantify the need for learning, we first computed the odds of a “random” readout to perform well. Random readout weights were drawn from a Gaussian distribution with the same mean and standard deviation as the optimal linear readout weights. We evaluated the performance of 1,000,000 such readouts per target pair used in the experiment assuming $\alpha = 0.1$ and saturating glomerular responses ($\lambda = 0$), always using the experimental trial structure (Figure 4A). The average performance was $58.41\% \pm 4.53\%$

(mean SD). The chance that a random readout performs better than 80% is around 1 in 10^5 . Thus, random hyperplanes typically fall short of separating the mixtures with target odors from the ones without. We next directly computed how many training samples are required to find the proper readout weights for performing the task. To find the minimum number of trials needed to generalize to the rest of the data, we varied the absolute number of training trials directly. With $\alpha = 0.1$, ~30 random trials were needed to get to 90% of asymptotic performance assuming linear mixing ($\lambda = 1$), and ~150 trials for saturating mixture responses ($\lambda = 0$) (Figure 4B). This difference highlights how saturation makes the task harder to learn, but crucially, these numbers of training trials compare favorably to mice.

The analysis above indicates that a linear decoder with random readout weights is unlikely to perform well, but it can learn to generalize from a rather small number of random mixture stimuli to thousands of other mixtures. Due to nonlinear mixing, this ability to generalize beyond trained samples strongly depends on the quality of training samples. A particularly meager training set is given by samples of just single odor components. Therefore, we trained classifiers only on single odors and tested them on arbitrary mixture stimuli. Without saturation, the OLE performance for mixtures with more than one component degrades only mildly but is substantially worse than an OLE trained with a training set containing all component numbers

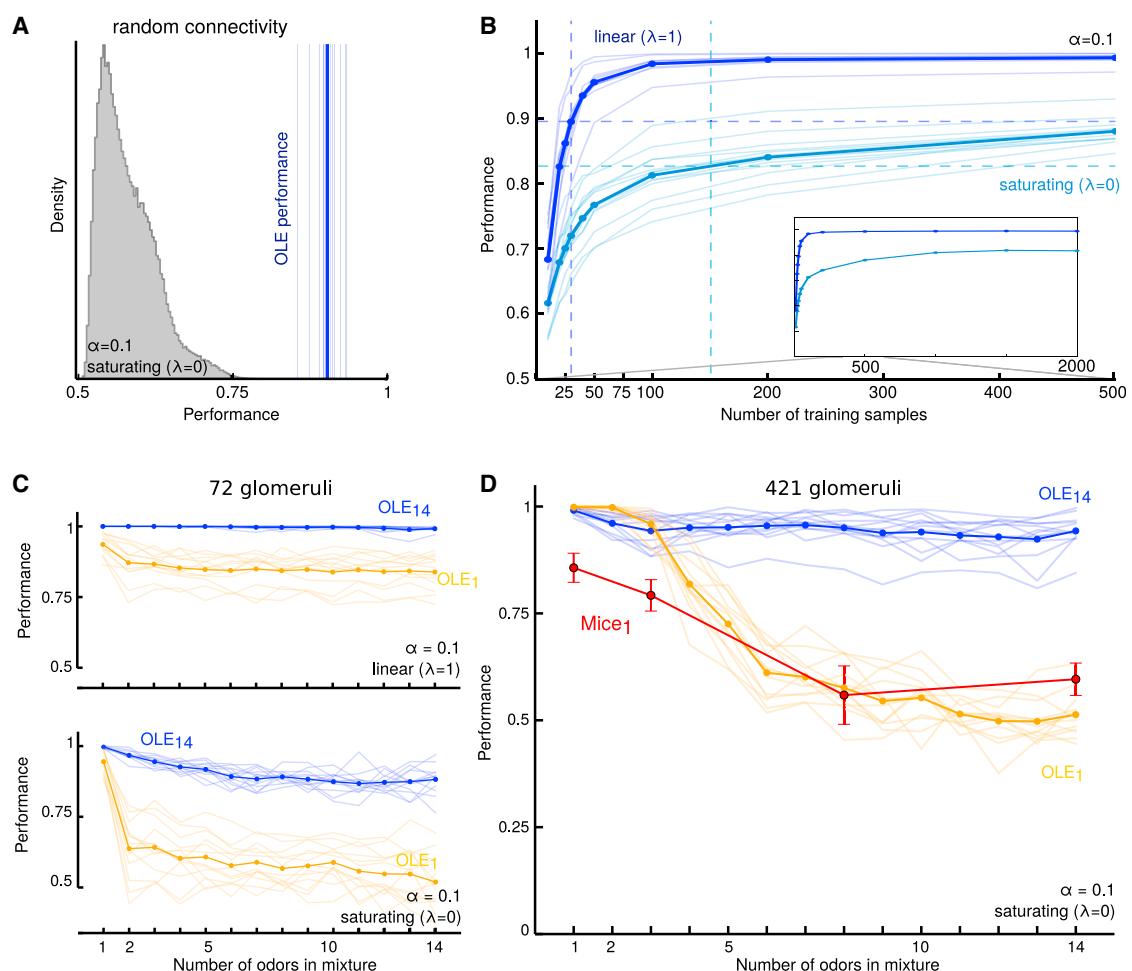


Figure 4. Learnability and Dependence on Trial Statistics

(A) Distribution of performances for 13,000,000 random linear decoders with weights drawn from the distribution of OLE weights and corresponding optimal threshold θ . The ability of 1,000,000 random readouts to detect the presence of the targets was evaluated for each of the 13 target odor pairs. The blue lines show the performance of the OLE.

(B) Performance versus absolute number of training samples for $\alpha = 0.1$ and $C = 10^6$. For each target, the average performance when trained on a small, random subset of samples is shown; each curve is averaged over 20 random training sets. The inset shows the asymptotic average performance, which is reached when trained by 1,500–2,000 trials. The dashed, horizontal lines indicate 90% of the asymptotic performance, and the corresponding vertical lines mark the necessary training samples to achieve this performance.

(C) Top: average performance curves for OLE per target pair trained only on 80% of the single odor stimuli (yellow, OLE₁) and on 80% of all the data (blue, OLE₁₄), with 0.1 noise and linear mixing based on 72 glomeruli. Bottom: same as top for non-linear mixing.

(D) Average performance curves for OLE per target pair when trained only on 80% of the single odor stimuli (yellow, OLE₁) and on 80% of all the data (blue, OLE₁₄) based on 421 glomeruli. The red curve shows the average performance \pm SEM of five mice trained on single odors and subsequently tested on mixtures of 1, 3, 8, and 14 odors (Mice₁).

(Figure 4C, top). With saturation, the performance of the OLE degrades strongly with an increasing number of odors and already has a performance of below 70% for mixtures of two components (Figure 4C, bottom). This inability to generalize is obviously stronger when fewer glomeruli are used for decoding. When we pool over all measured glomeruli (6 OBs, 421 dorsal glomeruli), the performance of the OLE decays more gracefully with the number of components (Figure 4D). Thus, for the linear readout, the mixture model, as well as the training stimuli, strongly affects the ability of the decoder to detect target odors.

Discriminative Learning versus Generative Modeling

An important idea that is complementary to the purely discriminative learning considered above is that neural representations are also formed in a task-independent, unsupervised way by learning a generative model of the sensory input (Mumford, 1994; Hinton and Sejnowski, 1999). The crucial advantage of generative modeling is the ability to learn tasks with a much smaller number of task-specific training samples by exploiting general knowledge about the nature of the world. For our demixing task specifically, this would mean that the glomerular representation (Figure 1B) is first demixed into a representation akin to

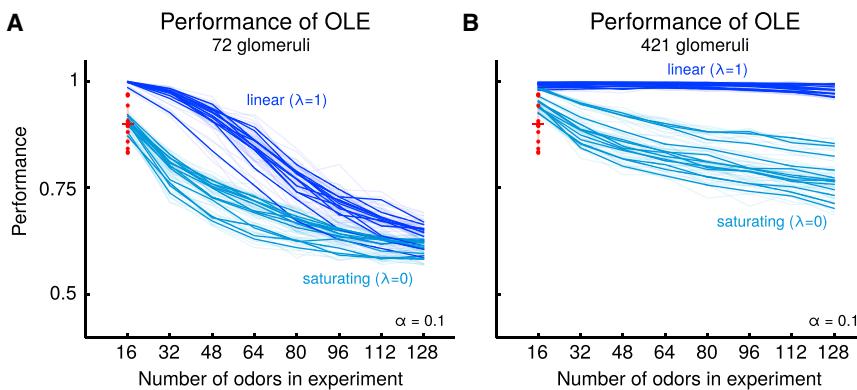


Figure 5. Capacity of an Optimal Linear Decoder in Cocktail Party Tasks with Increasing Odor Numbers

(A) Performance of optimal linear decoder with linear and nonlinear mixing and noise $\alpha = 0.1$ as a function of the number of odors. Additional odor patterns beyond 16 are generated by sampling from the responses of the 72 glomeruli to the 16 odors as described in the main text. Each faded line represents such a randomly generated odor dataset, and the 13 solid lines are the averages of four such lines with the same target odors. The red dots highlight the experimental performance of mice.

(B) Same as in (A), but calculated by pooling across imaging experiments to form odor patterns with 421 glomeruli.

the one in Figure 1A. Ideally this would allow the animal to perfectly learn the task from samples of just single odor components without the drop in performance observed for purely discriminative learning (cf. Figures 4C and 4D).

To test the ability of mice to generalize beyond single odor components, we trained five mice for several sessions on the single odor task until they reached >80% performance. During test sessions, we mainly presented single odors, but in 10% of the trials, we also presented mixture stimuli with 3, 8, and 14 components (equally distributed). Mice continued to perform well for single odorant trials and for (novel) mixtures of three. For mixtures with more components, performance dropped substantially (Figure 4D), consistent with OLE predictions. This pattern was observed in all mice individually (Figure S5C; Table S1).

Our earlier experiments demonstrated that mice can learn this task when exposed to training samples with more complex odor components (Rokni et al., 2014). This suggests that due to the nonlinear nature of odor mixture encoding, both mice and OLEs have difficulty generalizing much beyond the trained complexity, and rather settle on decision boundaries that are good enough to solve the task. It further indicates that mice cannot significantly benefit from a generative model by which they could task-independently decompose odor mixtures into their individual odor components.

Capacity of Linear Readout

Our analysis demonstrated that the presence of a single odor in mixtures of up to 14 odors is explicitly available for a linear decoder from the rich repertoire of receptors. What is the capacity limit for such a decoder? We can estimate this by using surrogate odor representations based on the representations of the 16 odors we measured. Surrogate odors were generated by drawing with replacement from the activations of the individual glomeruli by the 16 odors. We then calculated how well a linear decoder can report the presence of a pair of target odors in mixtures of up to 128 odors. Performance of classifiers was dependent on the target odor, but all classifiers performed above chance level even for 128 odors and above 80% for mixtures of around 32 odors (Figure 5A). This is remarkable since the classifier only uses 72 glomeruli out of a few thousand glomeruli in mice. We repeated the same analysis using pooled glomerular

data from six olfactory bulbs (421 glomeruli). With this large number of (potentially redundant) inputs, the performance of classifiers was above 70% for mixtures of up to 128 odors (Figure 5B). This analysis suggests that the capacity of mice to detect target odorants from mixtures may be well above what was tested.

DISCUSSION

An important task of the olfactory system is to identify odors of interest that are embedded in background mixtures. We formulated this task as a classification problem and asked whether a simple, biologically plausible classifier can solve this task in the face of noise and non-linear integration of odorants at the level of olfactory receptors. To tackle this question we first determined an encoding model for mixture responses of olfactory sensory neurons as the input for the classifiers. The model is based on data from calcium imaging and generates mixture representations that are saturating sums of single odor representations with trial-to-trial variability. We find that olfactory receptors have a substantial capacity to transmit information about the composition of odor mixtures despite saturation and noise, which is explicitly available for a linear, multiglomerular readout. In recent years, object recognition has been identified as a core challenge to understand the visual system (DiCarlo et al., 2012). For (visual) object recognition, simple linear readouts are also sufficient to match behavioral performance (DiCarlo et al., 2012; Majaj et al., 2015). However, these linear readouts cannot be applied directly to photoreceptors, but rather to the neuronal activity of higher visual areas, after several nonlinear transformation steps, which are generated from multiple feedforward steps of processing (DiCarlo et al., 2012). Our finding that the identity of odors in mixtures can be directly decoded by linear readouts from glomeruli highlights a fundamental contrast between the olfactory and visual systems, which is also mirrored anatomically: olfaction has substantially fewer anatomical stages than vision (Gire et al., 2013).

Mixture Model

The general question we have posed in this study is whether glomerular responses to mixtures that include target odors and those that exclude target odors are separable. Due to the large number of possible stimuli in the behavioral task (>10,000

possible mixtures), we estimated a statistical model of mixture responses based on measurements made from a sample of mixtures. We found that the relationship between glomerular responses to mixtures and the responses to mixture components can be described with a saturating nonlinearity. For the odor stimuli we used in this study, suppressive or supralinear responses were not significantly above response variability (Rospars et al., 2008). We also systematically measured the trial-to-trial variability in glomerular responses. We found two forms of response variability: one that is correlated across glomeruli and can probably be attributed to variation in the level of anesthesia and breathing and another that was uncorrelated across glomeruli and may represent trial-to-trial variability in individual input channels. The magnitude of the uncorrelated noise was, on average, 10% of response amplitude. In anesthetized rats, the CV of single OSN firing was found to be on the order of 1 (Duchamp-Viret et al., 2005). Single glomeruli receive a few thousand sensory axons, although this number can vary significantly depending on receptor gene (Bressel et al., 2016). If OSNs were statistically independent of each other, the glomerular CV is expected to be on the order of 1%–2%. The higher estimate in our study may indicate that sensory neuron responses co-vary and are not truly independent, but may also reflect non-biological sources of variability in our measurements.

Single Glomeruli Are Not Enough

In some cases, individual odorant receptors are thought to directly signal the presence of individual odors in a “labeled line” manner (Li and Liberles, 2015). These are thought to be involved in detecting specific odors that are important for innate behavior (Li and Liberles, 2015; Stowers et al., 2013). Here we find that individual glomeruli are not sufficient to report the presence or absence of individual odors. Although it is possible that our limited sample of dorsal glomeruli miss more specialized glomeruli with highly selective responses, it seems implausible that there will be such glomeruli for a vast array of odors. Indeed, experimental evidence strongly supports the idea that most odorant receptors in the main olfactory system are broadly tuned (Saito et al., 2009). Although individual glomeruli do not carry enough information to separate stimuli that contain a target odor from those that do not, we could reliably make this distinction by integrating information across different glomerular channels.

No Preprocessing Required; a Linear Readout Is Sufficient

Our decoders were based directly on the nonlinear, noisy glomerular inputs. Other than removing co-fluctuations across glomeruli, no intermediate stages of preprocessing such as normalization or decorrelation were necessary. Our analysis suggests that correlated trial-to-trial variability probably reflect slow changes due to anesthesia (Ecker et al., 2014) and breathing. Breathing-induced covariation could be easily removed by lateral interactions, which is why we neglected them here (Olsen et al., 2010; Blauvelt et al., 2013; Friedrich and Wiechert, 2014). Our measurements as well as those of others (Duchamp-Viret et al., 2005) indicate that noise levels at the glomeruli and overlap of glomerular patterns are not high enough to require an

additional processing stage. The optimal linear readout relies on weights that reflect the pattern of glomerular activation by target odors and are orthogonal to the pattern for distractor odors. This mechanism inherently removes overlap of glomerular patterns. Normalization and decorrelation may become more important for large amplitude stimuli that may saturate neuronal elements more strongly or for suboptimal readouts (like template matching). Decorrelation of different types has been proposed in the olfactory bulb, most notably in the context of reducing overlap among patterns of activity for different odors (Friedrich and Wiechert, 2014; Gschwend et al., 2015). Such pattern decorrelation or separation is thought to help downstream readout when noise limits separability (Friedrich and Wiechert, 2014). In the future, it will be interesting to look at the role and amount of noise correlations in glomeruli of awake animals performing this figure-ground segregation task.

Weights Can Be Sparse but Not Random

Although our intent was not to make specific analogies with neural circuits in the mouse brain, it is tempting to compare the decoder readout weights to the connections from the OB to the piriform cortex (PC). Since PC neurons receive a sample (possibly random) of glomerular input (Giessel and Datta, 2014), we also asked whether imposing sparseness constraints on the decoder affects performance. Remarkably, less than 20 glomeruli were sufficient to achieve performance that matches mice. It is likely that even fewer glomeruli may suffice if we were able to choose from the entire complement of mouse glomeruli. We found that while sparse connectivity was sufficient for robust performance after adequate training, classifiers built with random connectivity based on the same statistics as the optimal linear readouts without training were typically not successful. The problem remains linearly separable within the higher dimensional PC representation, but crucially, this recoding step is not necessary for a linear readout to be successful. Furthermore, our analysis of random readout suggests that even in a large PC population there might only be a few neurons that are selectively tuned to the target odors, due to the large number of mixtures. To the extent that we can make an analogy to mouse olfactory circuits, our findings suggest that any stochastic connectivity from the OB to PC may have to be supplemented with synaptic modification in the inputs, associative fibers, or outputs of the PC to allow odor mixture analysis. Whether such learning occurs associatively over an extended period as animals experience odor environments, or whether it requires more specific reinforcement, is an interesting question for future experiments. Alternatively, learning could happen via granule cells that dynamically gate the output of appropriate glomerular channels via mitral cells (Koulakov and Rinberg, 2011; Markopoulos et al., 2012), which would predict the existence of target-odor specific feedback modulation of mitral cells.

Although the optimal linear readouts are sufficient to match the performance of mice given the encoding model, we also showed that SVMs, which can approximate arbitrarily complex decision boundaries, outperform those simple decoders and suffer from decaying performance for increasing numbers of distractors. This performance decay suggests that for both mice and machines the ultimate bottleneck in this task is the overlapping

glomerular representation. While SVMs were not required to mimic mouse performance, it is conceivable that the myriad parallel pathways from the OB could implement decoding schemes similar to an SVM and that over time plasticity rules allow for the learning of intricate decision boundaries; however, for this segregation task, nonlinear decision boundaries are not necessary.

Robustness and Training Dependence

Other than the overall performance, the difficulty of classification can also be assessed by the robustness of classification to modifications in readout weights and by the speed at which a classifier converges on the proper weights. Even linear classifiers can learn the task relatively easily in dozens of trials, despite response variability and saturation. Although a direct comparison of learning rates of classifiers and mice may not be appropriate, it is worth noting that mice take hundreds of trials to learn the task under our conditions. We also find that once a classifier has been trained, small perturbations of the weights did not strongly affect performance. This indicates that the classifier has robust performance in a local region of the weight space, which nevertheless cannot be readily reached by random assignment. Our analysis of random readouts indicates that there is a 1 in 10^5 chance of having a performance of above 75%. Extrapolating this insight to neural architecture of the mouse olfactory system, it might mean that any projections that start out randomly (e.g., the OB to cortex projections) will need to be modified with learning, but that there might be several “template” neurons whose high correlation with reward will facilitate learning.

We showed that optimal linear readouts based on glomerular activity are highly sensitive to the quality of the training set. When trained on single odors, they perform poorly for mixtures. An ideal observer that knows how individual odors mix should perform accurately for mixtures even when trained on single odors. To be specific, assuming an explicit “atomic” representation of the single odors in the olfactory cortex, and that a decoder has access to this population, the readout should generalize well to mixtures when solely trained on the atoms. This reflects the main advantage of generative modeling, which has therefore been hypothesized to play an important role for shaping sensory representations (Barlow, 1997; Mumford, 1994; Hinton and Sejnowski, 1999). We tested this prediction in mice and found that while their performance for single odors remains stable and their performance for mixtures of three is comparable, mice generalize poorly to mixtures of 8 and 14. This suggests that mice learn the segregation task by adaptively adjusting decision boundaries directly based on incoming sensory information rather than highly processed, demixed representations. Due to the highly nonlinear representation of mixtures, such decision boundaries will generalize poorly when only trained with single odors. An alternative explanation for the mouse behavior (i.e., failure to detect target odor in mixtures when trained only on single odorants) is that they learned a different task rule—for instance, that 2 of the 16 odors are rewarded only when presented alone. Due to the novelty and low relative frequency of the mixture stimuli, they may have ignored those stimuli and thus failed to generalize. Such an interpretation, however, is at odds with the good performance for mixtures of three odors and the

fact that their errors consisted of both false alarms and misses. The training data from our previous study also argue against this alternate interpretation. The mice in that work were trained by picking odor mixtures with increasing complexities—from distributions ranging from few distractors to uniform distractor distributions. Whenever a mouse reached 80% performance, the distribution was changed (in three steps until the uniform distribution was presented in the test phase). There, the mice required greater than 1,000 trials to reach this level and usually also encountered hundreds of trials per condition. This experience dependency dwarfs the small number of trials that the linear decoder needs to learn the task. However, carefully designed future experiments are necessary to address the question of how much training is required for learning in mice. Arguably, learning the task relies at least on two independent components. First, mice have to learn the task structure (go/no go, reward, timing, etc.); second, mice have to learn which target odors are rewarded. By using the same task structure for different odor sets, one could greatly reduce the learning phase for the structure part and better estimate the number of trials that mice need for learning the task.

Experiments similar to those in the generalization task we studied in this paper (Figure 4D) have been done in humans. Jinks and Liang trained humans initially for 3 days on single chemicals and then tested them for mixtures (Jinks and Liang, 1999). The subjects detected single chemicals with performance of around 70%, and their performance for mixtures of 12 dropped to chance level. Our modeling, as well as our own experiments, suggests that this decay in performance is both a consequence of the training data and the synthetic encoding of stimuli, which makes generalization difficult. In vision, the intriguing characteristics of training schedules have long been noted—i.e., a minimal number of training examples is needed per session for learning, and an interleaved stimulus presentation design can hinder learning (Aberg and Herzog, 2012). Consequently, we believe that the study of the impact of training schedules could also be fruitful for understanding olfactory circuits in mice.

Capacity

Overlapping representations of odors allow the olfactory system to encode more odors than the number of receptor types (Hopfield, 1999). However, broad tuning of receptors may decrease the discriminability of stimuli when noise and saturating glomeruli are considered. An earlier analysis of the coding capacity of spatial patterns of glomerular activation (Koulakov et al., 2007) pointed out that even with all-or-nothing glomerular activity, it is possible to simultaneously represent a dozen mixed odorants. Our analysis considers graded glomerular responses and empirical measurements of responses to specific odors and suggests that the capacity of the system is substantially greater. Whether this estimate is comparable to the capacity of mice remains unknown. Yet, understanding the capacity to detect and discriminate odors (Bushdid et al., 2014; Meister, 2015; Weiss et al., 2012) will provide crucial insights for understanding the olfactory system.

Conclusions

In this work, we have linked experimentally measured glomerular responses to behavioral data in an odor demixing task. Using

realistic assumptions about neuronal noise, and nonlinear interactions for mixtures, we show that the information about odor-mixture components at the level of olfactory receptors is already linearly separable and does not require any preprocessing or inference algorithms that rely on prior information and feedback circuits.

EXPERIMENTAL PROCEDURES

All computational analyses are based on published behavioral and imaging data (Rokni et al., 2014). Further imaging experiments to estimate trial-to-trial variability and mixing properties in glomeruli, as well as additional behavioral experiments, have been performed (described below). All procedures were performed using approved protocols in accordance with institutional (Harvard University Institutional Animal Care and Use Committee) and national guidelines. Additional details for all the methods can be found in the [Supplemental Experimental Procedures](#).

Glomerular Imaging

Wide-field imaging was performed on adult OMP-GCaMP3 mice (Isogai et al., 2011) as described previously (Rokni et al., 2014) using a 10× objective (Olympus, NA 0.3). Blue light from an LED (M470L3, Thorlabs) was used for excitation, and the emitted light was filtered (500–550 nm MDF-GFP, Thorlabs) and collected with a CMOS camera (DMK 23U274, The Imaging Source GmbH). Odors were delivered using a custom-built automated olfactometer (Rokni et al., 2014). All odorants were diluted to 30% v/v in diethyl phthalate and then further diluted 16-fold in air. The 16 odors and their mixtures were presented in a random order for 2 s (inter-stimulus interval of 45 s).

Imaging Data Analysis

All odor responses were first converted into dF/F images with dF being the subtraction of the mean fluorescence in each pixel before odor presentation from the mean fluorescence during odor presentation. Putative glomeruli were identified from averaged dF/F images (across repetitions of the same; see [Supplemental Experimental Procedures](#)) and the time-varying dF/F trace for each glomerulus-odor pair. Response magnitude was quantified by integrating these dF/F traces.

Variability of single odor responses was measured as the coefficient of variation (CV) of glomerular responses. To remove correlated variability, we subtracted the best linear fit for each odor between the glomerular response at each trial O_j^t and the average response for that odor O_j from the individual trials O_j^t . We then calculated the remaining (private) variability for each glomerulus, i.e.:

$$\eta^t = O_j^t - (a^t O_j + b^t), \quad (\text{Equation 4})$$

where η^t is the deviation of glomerular response from the linear fit, O_j^t is the response to the j th odor on trial t , O_j is the trial-averaged response to odor j , and a^t and b^t are the slope and intercept of the best linear fit between O_j^t and O_j . For each glomerulus i , we then calculated the standard deviation of η^t and plotted it against its mean response amplitude O_{ij} . The non-correlated CV for each experiment was measured as the slope of the linear relationship between the non-correlated standard deviation and response amplitude (Figure 1H).

Summation of mixture components was analyzed by comparing mixture responses to the linear sum of the mean responses to mixture components. The data were fitted by a sigmoidal function (for each glomerulus and mixture):

$$\sigma(R_0) = \frac{2 \cdot A}{1 + e^{-R_0 \cdot s}} - A, \quad (\text{Equation 5})$$

where $\sigma(R_0)$ is the response to a mixture, A is the saturation level for the glomerulus, R_0 is the linear sum of mixture component responses, and s is a free parameter.

Mixture Models for Decoding Analysis

For the decoding analysis, we considered the following encoding model, which is based on our measurements. Given the average glomerular response

vectors O_j and the trial structure $c_j(t)$, i.e., a binary variable denoting the presence of odor j in trial t , the linear glomerular response pattern to a mixture in trial t is given by [Equation 2](#) in the main text. This mean response is subject to trial-to-trial variability. The linear response R_0 is thus the sum of the single component response O_j with an added noise term η_j that is normally distributed with a mean of zero and a variance of $\alpha^2 O_j^2$ as defined by [Equation 2](#). The factor α controls the coefficient of variation.

To account for saturation, we model the nonlinear response $\sigma(R_0)$ to a mixture in trial t as a sigmoid function of the linear response R_0 :

$$\sigma(R_0(t)) = \frac{2 \cdot A}{1 + e^{-R_0(t) \cdot s}} - A, \quad (\text{Equation 6})$$

where A is the saturated response amplitude and s is a free parameter that governs the slope of this sigmoidal function. For each glomerulus i , we assumed A to be equal to the maximum of the linear sum of component response $R_0^{(i)}(t)$ and set $s = 10/A$. This makes $\sigma(R_0^{(i)}(t))$ almost fully saturated for inputs larger than $A/2$, and the output level is around 50% of the saturation level for input values around $A/10$.

Decoding Analysis

The glomerular responses in a given trial t are drawn from the probability distribution for response vectors $R(t)$, which are generated using the mixture model. For any given trial t , the target was either present or not, and we denote this ideal behavior by the binary variable $r(t)$. Any readout function $f(R, \theta)$ will use $R(t)$ as its input, with parameters θ that need to be optimized. This optimization can be framed using the 0-1 loss function $\delta(r, f(R, \theta))$ and the empirical risk functional for a subset of trials I with cardinality $|I|$ ([Schoelkopf and Smola, 2002](#)):

$$A_{\text{emp}}(\theta) = \frac{1}{|I|} \sum_{t \in I} (1 - \delta(r(t), f(R(t), \theta))). \quad (\text{Equation 7})$$

Since the Kronecker- δ is one when both entries are the same, and zero otherwise, this empirical risk function measures the fraction of correct classifications by readout $f(R, \theta)$. The goal is then to find parameters θ_0 for the readout $f(R, \theta_0)$ that minimizes the empirical risk on the training set I . Since the 0-1 loss is not differentiable, and its optimization is NP hard, it is common to minimize surrogate loss functions that are convex and thus facilitate the determination of a unique optimum.

To compare different readouts, we compare their performance on the test set, i.e., the other trials that have not been used for training the readout. All read-outs are cross-validated using the repeated random subsample technique; unless otherwise specified, we used 80% training data, 20% test data, and 20 repetitions with randomly sampled training and test datasets. We report the average test performance, which is defined as the mean number of correct classifications of $r(t)$ by $f(R(t), \theta_0)$ averaged over the test set.

For the readout function f , we considered the following four options, which are commonly optimized based on different surrogate loss functions (details in [Supplemental Experimental Procedures](#)):

- (1) Single glomerulus readout: we used standard ROC analysis to obtain an upper bound for single glomerulus performance ([Figures S4C and S4D](#)).
- (2) Linear decoder: the linear decoder depends on the following parameters: the readout weights w and the decision threshold, and has output $y = H(w^T \cdot R(t) + \theta)$ ([Equation 3](#)). The optimal linear weights (OLE weights) are the ones that minimize the norm.

$$\|r(t) - w^T R(t)\|^2 \quad (\text{Equation 8})$$

All reported performances are the averages calculated by cross validation, which was used to determine the optimal threshold and weights. For each resampling, we calculated the OLE weights as described above and computed the optimal decision threshold θ as the one that maximizes the training performance.

- (3) Support vector machines: we studied the performance of SVMs with radial basis functions as kernels.

- (4) Logistic regression: we used regularized logistic regression to decode the presence of a target odor. For binary classification, the likelihood for the logistic regression is given by

$$\mathcal{L}_t = \rho(w^T R(t))^{r(t)} (1 - \rho(w^T R(t)))^{(1-r(t))}, \quad (\text{Equation 9})$$

with logistic function $\rho(s) = 1/(1 + \exp(-s))$. Minimizing misclassification can be formulated as minimizing the negative logarithm of the likelihood. Therefore, the error functional for the regularized logistic regression is given by

$$\sum_i |w_i| - C \sum_{t \in I} \log \mathcal{L}_t. \quad (\text{Equation 10})$$

We used the L1-norm to bias toward sparse read-outs, and the constant C varies the degree of sparseness as it balances the trade-off between classification errors and the summed absolute readout weights. We varied the regularization constant from 10^{-2} to 10^6 . We employed the Scikit-learn toolbox in Python to perform the SVM and logistic regression analysis (Pedregosa et al., 2011).

Details for the weight perturbation analysis, random readout analysis, and single-odor training simulations can be found in the [Supplemental Experimental Procedures](#).

Behavioral Testing

Behavioral training was performed as described earlier (Rokni et al., 2014). In brief, five C57BL/6 mice (Charles River) were trained on the task using single component odors. In each trial, a single odorant was presented for 2 s (intertrial interval of 10 s). One of two targets was presented in half the trials (go). Mice had to lick for go trials and refrain from licking for no go trials within the 2 s period. A correct "go" trial was rewarded with a water drop, correct rejections were not rewarded, and incorrect trials were punished by a 5 s timeout. In testing sessions, which began after mice performed above 80% correct in a session, mixtures of 3, 8, or 14 components (each mixture size equally probable) were presented every tenth trial.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.neuron.2016.08.007>.

AUTHOR CONTRIBUTIONS

A.M., D.R., M.B., and V.N.M. designed research; A.M. performed simulations and model analysis with input from M.B.; D.R. performed imaging and behavioral experiments; V.K. contributed the data on glomerular maps; and A.M., D.R., M.B., and V.N.M wrote the paper with input from all authors.

ACKNOWLEDGMENTS

We thank Philipp Berens, Mackenzie Amoroso, Gonzalo Otazu, and Alexandra Ding for helpful feedback. This work was supported by Harvard University, by DFG grant MA 6176/1-1 (A.M.), and by a Marie Curie Fellowship PIOF-GA-2013-622943 (A.M.). M.B. has received financial support from the Bernstein Center for Computational Neuroscience (FKZ 01GQ1002) and the German Excellence Initiative through the Centre for Integrative Neuroscience Tübingen (EXC307). Research in V.N.M.'s lab is supported by grants from the NIH (DC011291, DC014453). Computational resources were provided and maintained by Harvard FAS Research Computing. We thank the Kavli Institute for Theoretical Physics for hosting the authors at a workshop supported in part by the National Science Foundation grant PHY11-25915.

Received: November 17, 2015

Revised: May 31, 2016

Accepted: July 25, 2016

Published: September 1, 2016

REFERENCES

- Aberg, K.C., and Herzog, M.H. (2012). About similar characteristics of visual perceptual learning and LTP. *Vision Res.* 61, 100–106.
- Barlow, H.B. (1997). The knowledge used in vision and where it comes from. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 352, 1141–1147.
- Bell, A.J., and Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.
- Berens, P., Ecker, A.S., Cotton, R.J., Ma, W.J., Bethge, M., and Tolias, A.S. (2012). A fast and simple population code for orientation in primate V1. *J. Neurosci.* 32, 10618–10626.
- Blaauw, D.G., Sato, T.F., Wienisch, M., Knöpfel, T., and Murthy, V.N. (2013). Distinct spatiotemporal activity in principal neurons of the mouse olfactory bulb in anesthetized and awake states. *Front. Neural Circuits* 7, 46.
- Bressel, O.C., Khan, M., and Mombaerts, P. (2016). Linear correlation between the number of olfactory sensory neurons expressing a given mouse odorant receptor gene and the total volume of the corresponding glomeruli in the olfactory bulb. *J. Comp. Neurol.* 524, 199–209.
- Brody, C.D., and Hopfield, J.J. (2003). Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron* 37, 843–852.
- Bushdid, C., Magnasco, M.O., Vosshall, L.B., and Keller, A. (2014). Humans can discriminate more than 1 trillion olfactory stimuli. *Science* 343, 1370–1372.
- Cuevas Rivera, D., Bitzer, S., and Kiebel, S.J. (2015). Modelling Odor Decoding in the Antennal Lobe by Combining Sequential Firing Rate Models with Bayesian Inference. *PLoS Comput. Biol.* 11, e1004528.
- Del Castillo, J., and Katz, B. (1954). Quantal components of the end-plate potential. *J. Physiol.* 124, 560–573.
- DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434.
- Duchamp-Viret, P., Chaput, M.A., and Duchamp, A. (1999). Odor response properties of rat olfactory receptor neurons. *Science* 284, 2171–2174.
- Duchamp-Viret, P., Kostal, L., Chaput, M., Lánsky, P., and Rospars, J.-P. (2005). Patterns of spontaneous activity in single rat olfactory receptor neurons are different in normally breathing and tracheotomized animals. *J. Neurobiol.* 65, 97–114.
- Ecker, A.S., Berens, P., Cotton, R.J., Subramanyan, M., Denfield, G.H., Cadwell, C.R., Smirnakis, S.M., Bethge, M., and Tolias, A.S. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82, 235–248.
- Firestein, S., Picco, C., and Menini, A. (1993). The relation between stimulus and response in olfactory receptor cells of the tiger salamander. *J. Physiol.* 468, 1–10.
- Friedrich, R.W., and Wiechert, M.T. (2014). Neuronal circuits and computations: pattern decorrelation in the olfactory bulb. *FEBS Lett.* 588, 2504–2513.
- Galán, R.F., Weidert, M., Menzel, R., Herz, A.V.M., and Galizia, C.G. (2006). Sensory memory for odors is encoded in spontaneous correlated activity between olfactory glomeruli. *Neural Comput.* 18, 10–25.
- Giessel, A.J., and Datta, S.R. (2014). Olfactory maps, circuits and computations. *Curr. Opin. Neurobiol.* 24, 120–132.
- Gire, D.H., Restrepo, D., Sejnowski, T.J., Greer, C., De Carlos, J.A., and Lopez-Mascaraque, L. (2013). Temporal processing in the olfactory system: can we see a smell? *Neuron* 78, 416–432.
- Godfrey, P.A., Malnic, B., and Buck, L.B. (2004). The mouse olfactory receptor gene family. *Proc. Natl. Acad. Sci. USA* 101, 2156–2161.
- Gottfried, J.A. (2010). Central mechanisms of odour object perception. *Nat. Rev. Neurosci.* 11, 628–641.
- Grabska-Barwinska, A., Beck, J., Pouget, A., and Latham, P. (2013). Demixing odors – fast inference in olfaction. In *Advances in Neural Information Processing Systems, Volume 26*, C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, eds. (Curran Associates), pp. 1968–1976.

- Gschwend, O., Abraham, N.M., Lagier, S., Begnaud, F., Rodriguez, I., and Carleton, A. (2015). Neuronal pattern separation in the olfactory bulb improves odor discrimination learning. *Nat. Neurosci.* **18**, 1474–1482.
- Hinton, G., and Sejnowski, T.J. (1999). Unsupervised learning: foundations of neural computation (Cambridge, MA: MIT Press).
- Hiratani, N., and Fukai, T. (2015). Mixed signal learning by spike correlation propagation in feedback inhibitory circuits. *PLoS Comput. Biol.* **11**, e1004227.
- Hopfield, J.J. (1991). Olfactory computation and object perception. *Proc. Natl. Acad. Sci. USA* **88**, 6462–6466.
- Hopfield, J.J. (1999). Odor space and olfactory processing: collective algorithms and neural implementation. *Proc. Natl. Acad. Sci. USA* **96**, 12506–12511.
- Isogai, Y., Si, S., Pont-Lezica, L., Tan, T., Kapoor, V., Murthy, V.N., and Dulac, C. (2011). Molecular organization of vomeronasal chemoreception. *Nature* **478**, 241–245.
- Jinks, A., and Laing, D.G. (1999). A limit in the processing of components in odour mixtures. *Perception* **28**, 395–404.
- Koulakov, A.A., and Rinberg, D. (2011). Sparse incomplete representations: a potential role of olfactory granule cells. *Neuron* **72**, 124–136.
- Koulakov, A., Gelperin, A., and Rinberg, D. (2007). Olfactory coding with all-or-nothing glomeruli. *J. Neurophysiol.* **98**, 3134–3142.
- Li, Z. (1990). A model of olfactory adaptation and sensitivity enhancement in the olfactory bulb. *Biol. Cybern.* **62**, 349–361.
- Li, Q., and Liberles, S.D. (2015). Aversion and attraction through olfaction. *Curr. Biol.* **25**, R120–R129.
- Majaj, N.J., Hong, H., Solomon, E.A., and DiCarlo, J.J. (2015). Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. *J. Neurosci.* **35**, 13402–13418.
- Malnic, B., Hirono, J., Sato, T., and Buck, L.B. (1999). Combinatorial Receptor Codes for Odors. *Cell* **96**, 713–723.
- Markopoulos, F., Rokni, D., Gire, D.H., and Murthy, V.N. (2012). Functional properties of cortical feedback projections to the olfactory bulb. *Neuron* **76**, 1175–1188.
- Meister, M. (2015). On the dimensionality of odor space. *eLife* **4**, e07865.
- Mumford, D. (1994). Neural architectures for pattern-theoretic problems (Cambridge, MA: MIT Press).
- Olsen, S.R., Bhandawat, V., and Wilson, R.I. (2010). Divisive normalization in olfactory population codes. *Neuron* **66**, 287–299.
- Otazu, G.H., and Leibold, C. (2011). A corticothalamic circuit model for sound identification in complex scenes. *PLoS ONE* **6**, e24270.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.
- Rokni, D., Hemmeler, V., Kapoor, V., and Murthy, V.N. (2014). An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nat. Neurosci.* **17**, 1225–1232.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408.
- Rospars, J.-P., Lansky, P., Chaput, M., and Duchamp-Viret, P. (2008). Competitive and noncompetitive odorant interactions in the early neural coding of odorant mixtures. *J. Neurosci.* **28**, 2659–2666.
- Saito, H., Chi, Q., Zhuang, H., Matsunami, H., and Mainland, J.D. (2009). Odor coding by a Mammalian receptor repertoire. *Sci. Signal.* **2**, ra9.
- Schoelkopf, B., and Smola, A. (2002). Learning with kernels: support vector machines, regularization, optimization and beyond (Cambridge, MA: MIT Press).
- Shen, K., Toootonian, S., and Laurent, G. (2013). Encoding of mixtures in a simple olfactory system. *Neuron* **80**, 1246–1262.
- Stettler, D.D., and Axel, R. (2009). Representations of odor in the piriform cortex. *Neuron* **63**, 854–864.
- Stowers, L., Cameron, P., and Keller, J.A. (2013). Ominous odors: olfactory control of instinctive fear and aggression in mice. *Curr. Opin. Neurobiol.* **23**, 339–345.
- Toootonian, S., and Lengyel, M. (2014). A Dual Algorithm for Olfactory Computation in the Locust Brain. In *Advances in Neural Information Processing Systems, Volume 27*, Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, eds. (Curran Associates), pp. 2276–2284.
- Weiss, T., Snitz, K., Yablonka, A., Khan, R.M., Gafsiou, D., Schneidman, E., and Sobel, N. (2012). Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. *Proc. Natl. Acad. Sci. USA* **109**, 19959–19964.
- Wilson, D.A., and Sullivan, R.M. (2011). Cortical processing of odor objects. *Neuron* **72**, 506–519.