nature neuroscience

A probabilistic approach to demixing odors

Agnieszka Grabska-Barwińska[1,2], Simon Barthelmé[3], Jeff Beck[4], Zachary F Mainen[5], Alexandre Pouget[1,6–8] & Peter E Latham[1,8]

The olfactory system faces a hard problem: on the basis of noisy information from olfactory receptor neurons (the neurons that transduce chemicals to neural activity), it must figure out which odors are present in the world. Odors almost never occur in isolation, and different odors excite overlapping populations of olfactory receptor neurons, so the central challenge of the olfactory system is to demix its input. Because of noise and the large number of possible odors, demixing is fundamentally a probabilistic inference task. We propose that the early olfactory system uses approximate Bayesian inference to solve it. The computations involve a dynamical loop between the olfactory bulb and the piriform cortex, with cortex explaining incoming activity from the olfactory receptor neurons in terms of a mixture of odors. The model is compatible with known anatomy and physiology, including pattern decorrelation, and it performs better than other models at demixing odors.

The olfactory system has evolved to process information about chemicals in the environment. Much is known about the physiological side of this processing, especially in the early stages[1]. At the very first stage, neurons in the nasal epithelium, called olfactory receptor neurons (ORNs), transduce chemicals in the air into electrical signals. Each ORN expresses exactly one type of olfactory receptor, and in mammals there are about 1,000 different types. The question we address here is: how does the brain extract olfactory percepts from the ORN activity? More simply, how does it answer questions such as: given the relatively complex mixture of chemicals just inhaled, which odors are present? (We use "odor" to refer to the olfactory percept corresponding to a particular object, as in the "the odor of an orange," and "odorant" to refer to the many chemicals that are released by the orange.)

Inferring olfactory percepts from the ORN signals is difficult for several reasons. First, it is rarely the case that a single odor dominates the environment. Instead, multiple odors are typically present: in a restaurant there are many different dishes, in a forest many different plants, etc. Thus, the task of the olfactory system is more often segmentation than recognition, at least outside an experimental laboratory. And even when the task is to recognize a single odor, that typically must be done against a background of other odors[2]. Second, ORNs respond to a broad range of odors[3], so information about the olfactory scene is distributed across many neurons. Finally, because neural responses are stochastic, the same odor never elicits the same pattern of activity twice.

For all these reasons, olfaction is fundamentally a probabilistic inference task. We hypothesize that, when faced with an olfactory scene, the olfactory system computes a probability distribution over the possible odors. Although it is not known whether olfactory neurons encode probabilities, there is very strong evidence that other sensory modalities do[4], so this is a reasonable hypothesis. And there are good reasons to keep track of probabilities. For example, suppose you smell a fruit and conclude that there is an 80% chance it is a grapefruit and a 20% chance it is an orange. Suppose you then look at the fruit and, on the basis of the image, conclude that there is a 5% chance it is a grapefruit and a 95% chance it is an orange. Such cross-modal disagreement is easily resolved using the rules of probabilistic inference (assuming equal a priori probabilities, there is about a 17% chance it is a grapefruit). But it is only because the probabilities were known that an optimal decision—about, say, whether to eat the fruit—could be made. If the two modalities had returned binary answers ("it is a grapefruit" and "it is an orange"), there would be no principled way to resolve the conflict.

Here we present a model of how the early olfactory system—the olfactory bulb, along with the piriform cortex—could demix odors. More specifically, we show that the early olfactory system can, on the basis of a single sniff, compute a probability distribution over the concentration of each possible odor via a dynamic process involving the olfactory bulb and the piriform cortex. Previous work on demixing has either assumed that only one odor is detected on each sniff[5–7] or that different odors have different temporal patterns[8,9]. Our model both treats odors probabilistically and is capable of demixing multiple, temporally homogeneous odors in a single sniff.

RESULTS

A probabilistic model of olfaction

Essentially all probabilistic models of sensory processing proceed in three steps, and olfaction is no different: (i) specify an encoding model, a probabilistic mapping from odors to neural activity, (ii) specify a prior probability over odors, and (iii) use Bayes' theorem to invert the model and compute a probability distribution over odors given neural activity. Applying this procedure results in a set of equations that constitutes our inference algorithm—the algorithm for transforming neural activity into a probability distribution. In Online Methods

[1]Gatsby Computational Neuroscience Unit, University College London, London, UK. [2]Google DeepMind, London, UK. [3]CNRS, Gipsa-lab, Grenoble, France. [4]Department of Neurobiology, Duke University Medical School, Durham, North Carolina, USA. [5]Champalimaud Centre for the Unknown, Lisbon, Portugal. [6]Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland. [7]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York, USA. [8]These authors contributed equally to the work. Correspondence should be addressed to P.E.L. (pel@gatsby.ucl.ac.uk).

Received 13 May; accepted 21 October; published online 5 December 2016; doi:10.1038/nn.4444

section "Approximate inference," we provide a detailed description of the encoding model, the prior, and the method for inverting Bayes' theorem. Here we sketch the main ideas.

The encoding model specifies the activity of each ORN receptor type given the set of concentrations in the world. For that we use a relatively simple model: the firing rate of ORN receptor type i, denoted ν_i, is a weighted sum of the concentrations,

$$\nu_i = \nu_{0,i} + \Delta t^{-1} \sum_{j=1}^{K} w_{ij} c_j \qquad (1)$$

where c_j is taken to be log concentration, chosen so that $c_j = 0$ corresponds to a concentration so low that it is undetectable, $\nu_{0,i}$ is the background firing rate of ORN receptor type i, and Δt is the time window for counting spikes. We assume that the neurons fire with Poisson statistics, as has been observed, at least approximately[10]; thus, in our model the input to the olfactory bulb is a set of Poisson spike trains.

The second step of probabilistic modeling is to determine the prior probability distribution over concentration (the distribution over c_j). Here we make two assumptions. The first is that only a small number of odors are present at a time, meaning only a small number of the c_j are nonzero. The second is that odors occur independently of each other. While the former is reasonable (we rarely detect more than a handful of odors at any one time) the latter is not correct. That's because odors tend to be correlated; for example, the set of odors one expects in a restaurant are different from those one expects in a forest. However, modeling those dependencies would require a complex, hierarchical prior. While such a prior is, ultimately, important for a complete understanding of olfaction, it adds complexity without changing the basic story; we thus leave it for future work.

We can satisfy both assumptions by combining, for each odor, a smooth function, which describes concentrations above detection threshold, with a point distribution whose total probability corresponds to the fraction of time concentrations are so low as to be undetectable. This gives us a prior of the form

$$p(\mathbf{c}) = \prod_j \left[(1 - p_{\text{prior}})\delta(c_j) + p_{\text{prior}} \frac{e^{-c_j/\beta_{\text{prior}}}}{\beta_{\text{prior}}} \right] \qquad (2)$$

where p_{prior} is the prior probability that any particular odor is present, β_{prior} sets the characteristic scale of the concentrations, and $\delta(c_j)$ is a delta function (the point mass). In our simulations we use $\beta_{\text{prior}} = 3$ and $p_{\text{prior}} = 3/K$ where $K = 640$ is the number of odors; the latter implies that there are, on average, 3 odors present in any particular olfactory scene. The precise shape of the distribution is not so important. What is important is that the expected number of odors is small, so that the olfactory system infers mixtures with a small number of odors.

The third step is to invert the generative model—that is, combine the prior (equation (2)) with the encoding model (equation (1))—to determine the probability that any particular odor is present. To do that, the olfactory system needs to know the possible odors in the world and the set of weights, w_{ij}, that transform those odors to ORN activity. In a full treatment, these would be learned; here we assume learning has already occurred. Or, more precisely, we assume that a subset of the possible odors, and their corresponding weights, have been learned. We refer to the learned subset as "known" odors. All other odors are termed "unknown."

Exact inference in this model is not feasible, as computation time is exponential in the number of odors; this is typical of Bayesian inference problems in realistic settings. We therefore use an approximate algorithm, which, as we will see, can be implemented in neural circuitry consistent with known anatomy and physiology[11,12]. The algorithm finds the factorized distribution that is as close as possible to the true one. As we show in Online Methods section "Approximate inference" (in particular, equation (21a)), this results in a posterior distribution for the concentration of each odor. The posterior for odor j has the form

$$q_j(c_j \mid \mathbf{r}) \propto c_j^{\bar{c}_j/\beta_j - 1} e^{-c_j/\beta_j} \qquad (3)$$

where $\mathbf{r}$ ($\equiv r_1, r_1, \ldots$) is a vector of spike counts, with $r_i(t)$ set to the number of spikes from ORN receptor type i in a window of size Δt; in our simulations, we set Δt to 50 ms. The parameters $\bar{c}_j$ and β_j determine the approximate probability distribution over the concentration of odor j. These parameters have a natural interpretation: $\bar{c}_j$ is the mean concentration and $\bar{c}_j\beta_j$ is the variance. Both are important for inferring whether or not an odor is present. In particular, the lower the mean the less likely an odor is to be present, and for a fixed mean, the higher the variance the more likely the odor is to be present because it is more likely that the true concentration is relatively high.

Note that we cannot represent arbitrary posterior distributions over concentration; instead, our posterior is summarized by two parameters. Moreover, although both are important, the second one, β_j, turns out to be independent of activity, $\mathbf{r}$, and only weakly dependent on odor, j (see Online Methods, equations (22b) and (38)). Thus, the distribution in equation (3) is reasonably well summarized by $\bar{c}_j$, and that is what we focus on. Not surprisingly, given the difficulty of the inference task, $\bar{c}_j$ depends on the activity of the ORNs, $\mathbf{r}$, in a complicated way. However, the $\bar{c}_j$ can be computed by the network shown in **Figure 1**. Explicit equations describing the time evolution of the various cells in the network are given in Online Methods, equation (28). Here we provide a qualitative description.

Input to our network comes from the ORNs: each ORN receptor type projects, via a glomerulus (which we do not model), to one mitral cell; those cells then interact, via approximately reciprocal dendrodendritic connections, with the granule cells. We ignore any spread of signals in the granule cells; essentially, each mitral cell is considered to have its own private connection. However, as is observed experimentally[13], activity at the soma of the granule cells is transmitted to its dendrites and modulates activity there. The mitral cells also project, via the lateral olfactory tract, to the 'mean concentration' cells in piriform cortex—the cells labeled $\bar{c}_j$ in **Figure 1**. The mean concentration cells then project back to the granule cells, which in turn inhibit the mitral cells.

As this explanation suggests, the network implements a negative feedback loop: the mitral cells excite the mean concentration cells in piriform cortex, those cells feed back to the bulb and excite the granule cells, and the granule cells inhibit the mitral cells. That feedback loop acts iteratively: it infers a set of odors, compares that inference to the incoming information, uses the comparison to refine the inference, and then repeats the process. Here we illustrate this for an olfactory scene containing, for definiteness, three odors.

To simplify the analysis, we focus on mitral cells and mean concentration cells. The granule cells are of course critical to the operation of the network, but as shown in Online Methods (in particular, equation (31)), their main effect is to provide divisive inhibition. For our

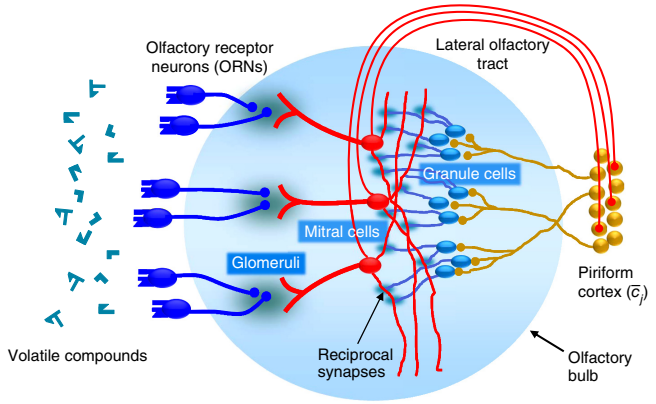

Figure 1 Circuit diagram. The olfactory receptor neurons (ORNs), which respond to volatile compounds in the air, are divided into different receptor types. All ORNs of the same receptor type project, via a glomerulus (which we do not model), onto a mitral cell. (Each of our mitral cells should be thought of as a 'meta-mitral' cell, since in the real circuit there are approximately 15 mitral cells for every glomerulus.) The mitral cells interact, via approximately reciprocal connections, with the granule cells (which are inhibitory and, in our network, outnumber the mitral cells by a factor of 3). The mitral cells also project, via the lateral olfactory tract, to the mean concentration cells in piriform cortex (labeled $\bar{c}_j$). These are the cells that carry direct information about the distribution over concentration for odor j. They also provide feedback—essentially an estimate of the mean concentration—to the granule cells: when the odors are successfully inferred, the feedback signal cancels, via the inhibitory granule cells, the feedforward drive from the ORNs, and the mitral cells are pushed toward their baseline firing rates.

network the specific form of the divisive inhibition involves a square root: the activity of the ith mitral cell, m_i, is given approximately by

$$m_i \propto \frac{(\gamma_i r_i)^{1/2}}{\left(v_{0,i} + \Delta t^{-1} \sum_j w_{ij} \bar{c}_j\right)^{1/2}} \qquad (4)$$

where r_i is, as above, the activity of ORN receptor type i and γ_i is an unimportant scale factor (it cancels the factor of γ_i^{-1} that appears in equation (5) below). The denominator in this equation is due to the granule cell feedback, which provides targeted divisive inhibition via the w_{ij}—the same weights that generated the ORN responses (see equation (1)).

The mitral cells drive the mean concentration cells, thus completing the feedback loop. As shown in Online Methods (in particular, equation (28a)), the drive to mean concentration cell i is linear in m_i^2,

$$\tau_c \frac{d\bar{c}_j}{dt} + \bar{c}_j - \alpha_0 \beta_j \propto \bar{c}_j \sum_i \gamma_i^{-1} m_i^2 w_{ij} \qquad (5)$$

where τ_c is the time constant of the mean concentration cells and $\alpha_0 \beta_j$ is the mean concentration associated with the prior. The main effect of the term $\tau d\bar{c}_j/dt$ is to introduce a delay, as it forces the mean concentration cells to integrate the drive from the mitral cells (a similar term should appear in equation (4), but it is less important, so to simplify the explanation we did not include it). Note that the mitral cells drive the mean concentration cells via the transpose of the weights.

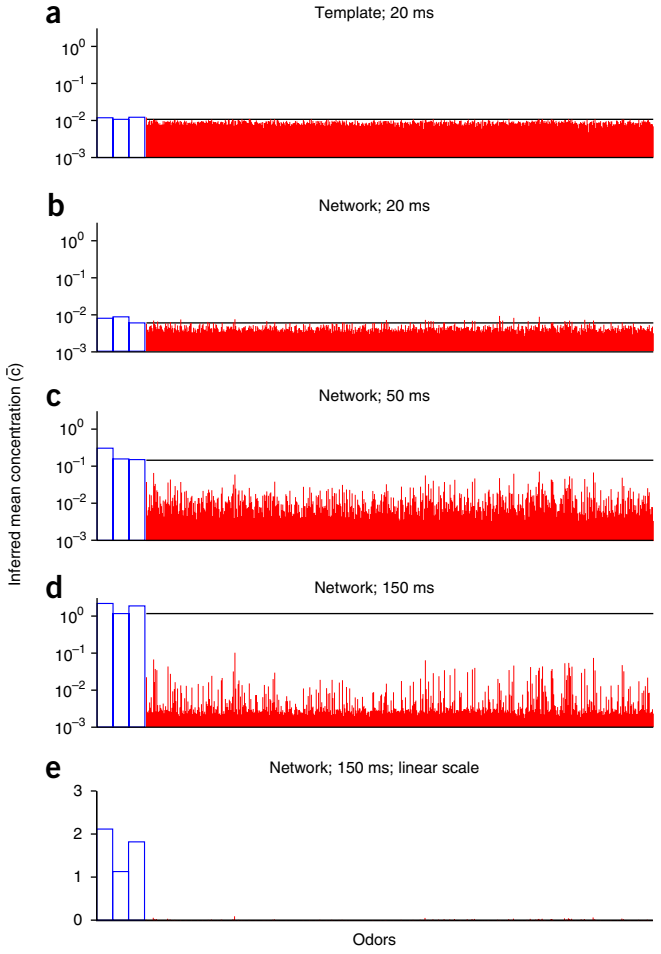

Figure 2 Evolution of activity in the mean concentration cells (the $\bar{c}_j$) when three odors are presented. The first three odors (blue) correspond to the presented odors; the remaining 637 (red) correspond to the odors that were not presented. All plots except that in **e** are on a log scale. (**a**) Template matching estimate, $\sum_i r_i w_{ij}$, 20 ms after odor onset. (**b**) Activity of the mean concentration cells 20 ms after odor onset. (**c**) Activity 50 ms after odor onset. (**d**) Activity 150 ms after odor onset. (**e**) Activity 150 ms after odor onsets, but on a linear, rather than a log, scale.

Immediately after odor onset, $\bar{c}_j$ is small—it is not far from $\alpha_0 \beta_j$ (which is about 0.002)—and so at early times the second term in the denominator of equation (4) can be ignored. Consequently, $m_i \propto r_i^{1/2}$, and mitral cells closely track ORN activity (see the section "Response to a known odor" below). This close tracking does not last long, though, only about 20 ms. However, that 20 ms is critical: replacing m_i^2 by $\gamma_i r_i$ in equation (5), we see that the mean concentration cells in piriform cortex are driven by $\Sigma_i r_i w_{ij}$. This is an approximation to the template matching signal (see **Fig. 2** and Online Methods section "ROC curves"), and in the first 20 ms it provides the network's initial estimate of which odors are present. The quality of that estimate can be seen in **Figure 2b**, which shows the activity of the mean concentration cells at 20 ms. Even though only 20 ms have elapsed, the presented odors are mainly larger than the non-presented ones. More quantitatively, a threshold set to the smallest presented odor would produce only 21 false positives (out of 637 possible false-positives). This is much better than chance, indicating that the initial feedforward sweep of activity is beneficial. The activity at 20 ms is also qualitatively similar to the actual template matching signal (**Fig. 2a**),

in the sense that a threshold set to the smallest presented odor would produce 15 false positives. (The lack of quantitative similarity between **Fig. 2a** and **b** is due primarily to nonlinearities in the network.)

Once inferred odor concentrations start to grow, their growth accelerates. That occurs because of the factor of $\bar{c}_j$ on the right hand side of equation (5). This creates a positive feedback loop and thus leads to a 'rich get richer' effect. But the rich do not get uniformly richer; there is also a consistency requirement. Because of the negative feedback, the mean concentration cells need appropriate drive from the ORNs to sustain their growth. This is a collective phenomenon, which can be understood by considering the following example. Suppose w_{i1} is nonzero for $i = 1, \ldots, 10$ and zero for all the other i, so odor 1 activates mitral cells 1–10. If only odor 1 were present, ORNs 1–10 would be activated (see equation (1)), and they would drive mitral cells 1–10 (see equation (4)). Those ten mitral cells would then strongly drive the cell representing odor 1 (see equation (5)). They would, of course, also drive other cells, but less effectively, as only a fraction of the ten active mitral cells, not all of them, would affect the other cells. Because of the negative feedback, the cells representing the other odors would not be able to sustain their activity. Thus, $\bar{c}_1$ would grow, but the other inferred odor concentrations would stay near background.

In more realistic situations the idea is the same: an initial template matching signal causes the correct odor to have reasonably high activity; the rich-get-richer effect (the factor of $\bar{c}_j$ on the right hand side of equation (5)) causes odors with initially elevated activity to grow; and consistency ensures that the correct odors are the ones most likely to eventually reach relatively high activity. Note that in our network we use a prior that does not favor any particular odor, but it is easy to change the prior in an odor-specific way simply by letting p_{prior} in equation (5) depend on odor, j; that is, let $p_{\text{prior}} \rightarrow p_{\text{prior},j}$, with $p_{\text{prior},j}$ larger for odors that are more likely to appear.

The growth of the odors is illustrated in **Figure 2**. We have already seen that the activity at 20 ms (**Fig. 2b**) provides a reasonable estimate of which odors are present. Only 50 ms later (**Fig. 2c**), the presented odors are starting to rise above the noise. In fact, a threshold set to the activity level of the smallest presented odor is a factor of more than 2 larger than the activity level of the largest non-presented odor. And at 150 ms (**Fig. 2d**,**e**), the smallest presented odor is an order of magnitude larger than all of the non-presented odors.

Although the correct odors were inferred, their mean concentrations were underestimated: the presented odors all had concentrations of 3, but the inferred mean concentrations were between 1 and 2 (**Fig. 2e**). This is a typical side effect of Bayesian inference, in which inferred quantities are biased toward the prior[14], which in our case favors low concentrations.

In this example the activity of the mean concentration cells was eventually either very high or very low, so it was clear which odors were present and which were not. However, had the activity been at an intermediate level it would have been less clear. To handle those cases, we need a mapping from the approximate posterior distribution (equation (3)) to the probability that an odor is present. Because, as discussed above, the approximate distribution is characterized primarily by the mean concentration, $\bar{c}_j$, we focus on the mapping from $\bar{c}_j$ to probability. To get that mapping, we performed many simulations, each with a different combination of odors, and combined them to compute the probability that an odor is present given $\bar{c}_j$ (see Online Methods section "Determining the probability that an odor is present"). This yielded a plot of probability versus $\bar{c}_j$ (**Fig. 3a**). Animals can easily learn, from experience, an approximation to this relationship. In addition, the inferred mean concentration, $\bar{c}_j$,

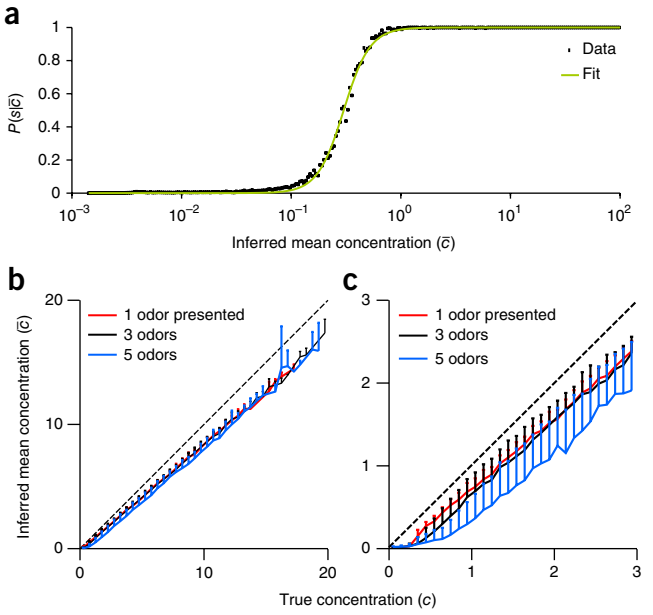

Figure 3 Probability of odor presence and inferred concentration. (**a**) Probability that an odor is present given the inferred mean concentration, $\bar{c}_j$, computed by the network from 10,000 trials. (**b**) Inferred mean concentration ($\bar{c}_j$) versus true concentration (c_j); error bars (shown only in the positive direction to reduce clutter) are s.d. On this scale, the inferred mean concentration is approximately proportional to (but slightly smaller than) the true concentration. (**c**) Same as **b**, but for concentrations between 0 and 3. Especially at low concentrations, our network underestimates the true concentration, with the underestimate larger when more odors are presented. This is due to the sparse prior on odors, which introduces a bias toward low inferred concentration. All three panels were constructed by performing a large number of simulations, each with a different olfactory scene; see Online Methods section "Determining the probability that an odor is present." In **b** the average number of data points per bin was 37, 170 and 127 for 1, 3 and 5 presented odors, respectively, and in **c** the corresponding numbers were 14, 31 and 144.

is well predicted, but slightly lower than, the true concentration, c_j (**Fig. 3b**,**c**). Thus, above a concentration of about 0.3–0.5 (depending on the number of presented odors) the inferred mean concentration, and thus the probability than an odor is present, is invariant to concentration.

Compatibility with known physiology

Our circuit is, of course, simplified relative to the true one. In the true circuit, each ORN receptor type projects to two glomeruli (one on either side of the olfactory bulb); these in turn innervate mitral cells, with each glomerulus innervating about 15 mitral cells; and each mitral cell forms dendro-dendritic connections with about 30 granule cells[15]. Signals can propagate within granule cells and so inhibit other mitral cells. In addition to this main circuitry, there is lateral inhibition among the glomeruli[15]. Finally, tufted and mitral cells have different response properties and projection patterns[16,17], and connectivity to and from the olfactory bulb is not limited to the piriform cortex[18]. In sum, we have bypassed the glomeruli, replaced the set of mitral and tufted cells by one meta-mitral cell, reduced the number of granule cells, and ignored the spread of depolarization in those cells. It will be important to include this more complex circuitry in future work. However, our circuit does contain the main cell types, mitral cells and granule cells, it preserves the approximately reciprocal

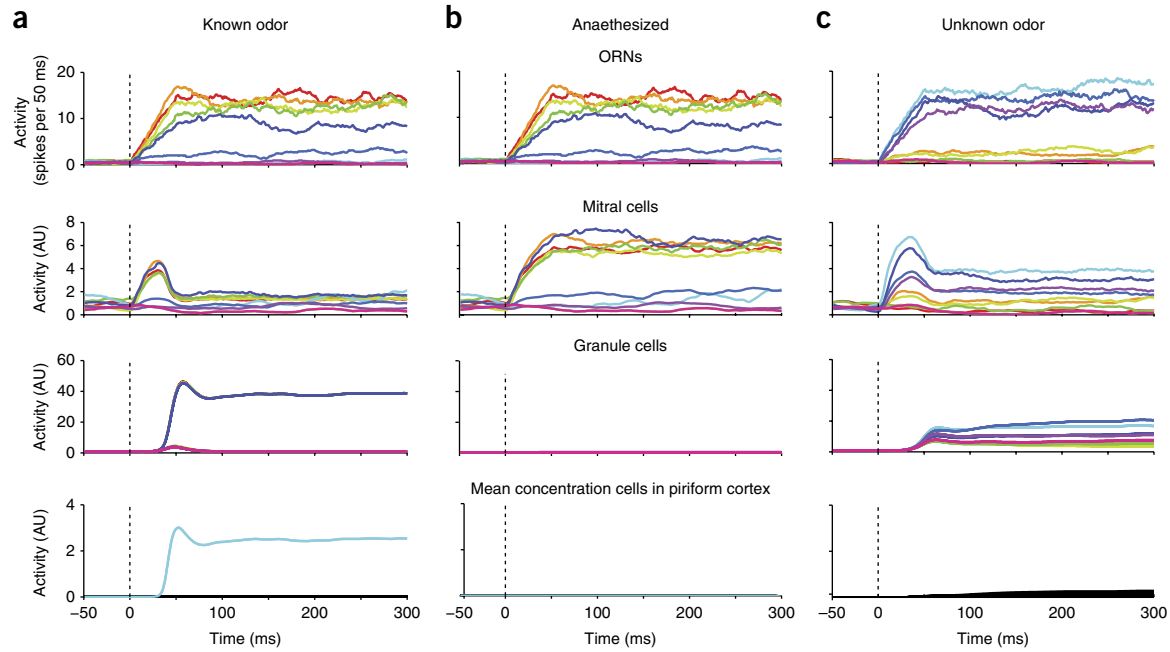

Figure 4 Activity in response to a single odor whose concentration is equal to 3, the mean concentration of present odors. Top to bottom: activity of 8 ORN receptor types; activity of the 8 mitral cells that receive input from the ORNs in the top row, matched by color; activity of the 8 granule cells whose main input comes from the mitral cells in the second row, matched by color; activity of all 640 mean concentration cells in piriform cortex. ORNs: activity is the number of spikes in 50 ms. Other cells: activity is in arbitrary units (AU). The traces are an average over 10 trials. (**a**) The odor is known to the animal. The cyan line in the bottom panel corresponds to the presented odor (odor 1), the (barely visible) black lines correspond to all the other odors. (**b**) The odor is known to the animal, but the feedback from the mitral cells to piriform cortex is reduced by a factor of 100. Because there is almost no cortical activity, granule cells lose a large component of their excitatory input, and so have low activity—too low and indiscriminate to cancel mitral cell activity. (**c**) The odor is unknown to the animal. The cortex tries to explain the ORN activity with many odors at low concentration, resulting in indiscriminate granule cell activity and only partial cancellation of mitral cell activity. There is no cyan line, as the presented odor has not been learned.

connections between those cell types, and it faithfully mirrors the main connections between the olfactory bulb and piriform cortex.

In the next several sections we investigate how this circuit responds to a range of olfactory scenes. For all simulations the network contained 160 ORN receptor types, 160 mitral cells, 480 granule cells, and 640 piriform cortex mean concentration cells, and there were 640 possible odors (see Online Methods sections "Approximate inference" and "Parameters" for parameters and details).

Response to a known odor

We first tested our model in response to a single known odor. Such an odor was generated by setting the concentration of odor 1, c_1, to 3 (the mean value of presented odors; see equation (2) and note that $\beta_{\text{prior}} = 3$) and the concentrations of all other odors to zero. Activity of our four cell types—ORNs, mitral cells, granule cells and mean concentration cells (the cells in piriform cortex that read out the odor)—are shown in **Figure 4a**. The ORN activity quickly rises to a steady state and stays there. The mitral cells, however, have more interesting dynamics: they exhibit an initial burst of activity from the ORN input, but that activity is terminated by inhibitory feedback from piriform cortex. In this example, the circuit correctly inferred which odor was present, so only one of the mean concentration cells (the correct one, $\bar{c}_1$) had appreciable activity.

The key observation is that when the odor is correctly inferred, activity in cortex mainly cancels the incoming ORN signal, and so mitral cell activity drops (as also predicted by Koulakov and Rinberg[19]). This is broadly consistent with data in awake, behaving animals[20–23]. To illustrate the importance of this cancellation, we 'anesthetized' cortex: we reduced the strength of the feedback connections from the mitral cells to the mean concentration cells by a factor of 100. With activity in cortex nearly nonexistent, granule cells have relatively low activity (**Fig. 4b**). Thus, mitral cell activity stays high, as it mainly represents the raw ORN signal. This is what is observed in rabbits when fibers from piriform cortex to the bulb are cut[24] and in mice when piriform cortex is pharmacologically suppressed[25].

Response to an unknown odor

A milder version of this lack of cancellation can be seen when an unknown odor is presented; we mimic this by setting the mean firing rate of ORN receptor type i to $\nu_{0,i} + \tilde{w}_i / \Delta t$ where the $\tilde{w}_i$ are a set of weights that do not correspond to any of the odors known to the animal. Because $\tilde{w}_i$ is not known by the circuit, it cannot infer which odor is present. Instead, piriform cortex tries to explain the ORN activity with many odors at low concentration (**Fig. 4c**). The activity in piriform cortex is sufficient to elicit granule cell activity; however, it is not the right activity, in the sense that it does not cancel the ORN input. Thus, although mitral cell activity is slightly lower than in **Figure 4b**, it does not return to the values seen in **Figure 4a**.

Detecting an odor against a background

One of the most common tasks faced by an animal is to detect a known odor against a background consisting of many other odors. Humans can do this reasonably well: they can detect a familiar odor in a background of 12 odors with about 65% accuracy[26]. And, as shown by Rokni *et al.*[2], mice can do even better: they can determine, with over 80% accuracy, whether or not a target odor is present against a background containing up to 13 distractor odors. We assessed the performance of our model on this task using a protocol identical to

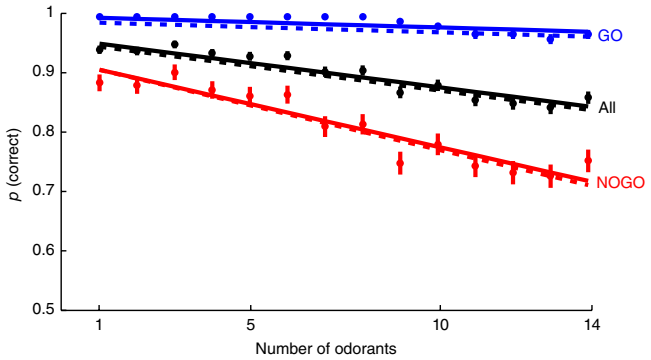

Figure 5 Performance of the model when detecting a known odor (target) against background odors. The target was present with probability 1/2. The *y* axis is the fraction of correct trials as a function of the number of odors present. Data are from our simulations, solid lines are least square fits, and dashed lines are fits to the experimental data of Rokni *et al.*[2]. Black: probability correct for all trials. Blue: probability correct for Go trials (trials containing the target). Red: probability correct for NoGo trials (trials not containing the target). For the Go and NoGo trials, each point corresponds to 500 simulations, with each simulation consisting of a randomly chosen set of odors. Error bars are s.e.m. assuming binomial statistics: error bars are $\pm(p(1-p)/n)^{1/2}$, where p is the probability correct and $n = 500$ is the number of simulations per data point.

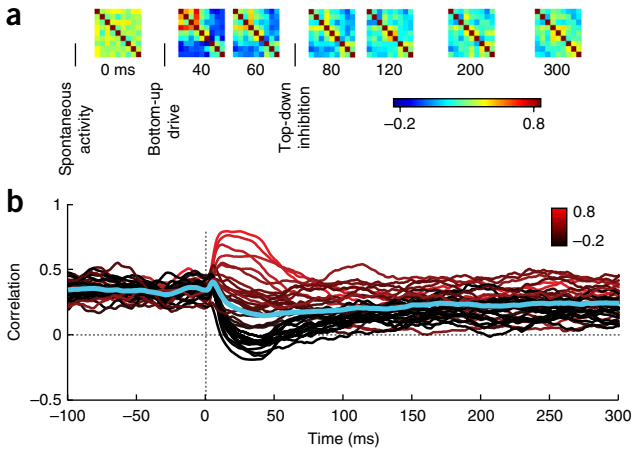

Figure 6 Evolution of pattern correlations in our model. (**a**) Pattern correlation matrices (Pearson correlation coefficients) for mitral cells at several time points. At time $t = 0$, before odor onset, activity is independent of odor; correlation coefficients are below 1 only because they were computed from a finite number of trials. In the early, 'bottom-up' phase ($t = 40$ ms), correlations reflect odor similarity. However, in the later, 'top-down' phase ($t \geq 80$ ms), feedback from the cortex forces a return to baseline, which pushes correlations down slightly. Decorrelation occurs because of this shift, but only on average; there are many pairs that increase in correlation. Color indicates correlation coefficient. (**b**) Same data as in **a**, but plotted for all pairs of odors versus time. Color (black to red) indicates input pattern correlations (correlations among ORN responses); blue is average correlation. The average correlation after odor onset was statistically significantly different from that before odor onset ($P < 10^{-6}$, two-sided *t*-test at −50 and 250 ms; $n = 36$, test statistic $t = 5.7$).

that in the experiments of Rokni *et al.*[2]. Briefly, we presented between 0 and 14 background odors, with half the trials also containing a known target odor. The concentrations of all odors, both target and distractors, was the same, $c = 3$. We reported the target present if its mean concentration ($\bar{c}_j$) was above a threshold of 0.076 and absent if it was below that threshold. In addition, we assumed that the mice licked spontaneously on 11% of the NoGo trials (the trials for which the animals are supposed to refrain from licking). The value of the threshold for our network, and the spontaneous lick rate on NoGo trials, were chosen to match as closely as possible the experimental results of Rokni *et al.*[2].

Our simulations match well the experimental results (**Fig. 5**), especially considering that we did not specifically train our network on this task. However, this task may be sufficiently simple that it is not a strong test of our Bayesian algorithm: it turns out that a linear decoder applied to the glomeruli activity also matches the behavior of the mice reasonably well[27]. We return to this point in the Discussion.

Pattern decorrelation as a side effect of inference

In a now seminal study, Friedrich and Laurent[28] showed that average mitral cell activity in the decerebrate zebrafish decorrelates over time: the average activity associated with two odors can be very similar (highly correlated) at the beginning of a trial, but very different (uncorrelated) at the end of a trial. (Note that we are referring to so-called pattern correlation, which measures the similarity of the average responses to a pair of odors.) The conclusion was that the olfactory bulb actively separates odors, making them easier to decode as time goes on[28].

To determine whether our model exhibits a similar pattern of decorrelation, we measured the average mitral cell responses to nine randomly chosen odors, presented one at a time, and computed, from ten trials, the correlation coefficient between each pair of odors. This gave us a time series of 9 × 9 correlation matrices (**Fig. 6a**). Some of the odors are initially strongly correlated, but over time their correlations weaken. Thus, odors that are initially very similar become less similar as time goes on—exactly what was seen by Friedrich and Laurent[28]. However, other odors become more correlated with time, as also seen by Friedrich and Laurent[28]. The net trend is that after odor onset the correlations decrease slightly (**Fig. 6b**). This decrease was recently observed in mouse olfactory bulb mitral cells[23], where, as in our simulations, the correlations dropped from about 0.4 to about 0.2.

Critically, decorrelation in our model does not imply that odors are easier to distinguish as time goes on. In fact, the opposite is true: mitral cells contain less information, not more, as time goes on. That is because during inference, granule cells act to cancel the ORN input to the mitral cells. This is consistent with what has been reported in awake mice[21,29]: decoding performance based on mitral cell activity increased immediately after odor onset, owing to the arrival of information in the olfactory bulb, but decreased a short time later. It is inconsistent, however, with the study of Friedrich and Laurent[28] in zebrafish, who found that decoding performance improved continually. This may be because the latter experiments were performed in decerebrate fish, where cortical feedback does not operate normally. Moreover, the decorrelation took place over several seconds, whereas in rodents a single breathing cycle (200–300 ms) is often sufficient for odor discrimination[30]. It is therefore unclear whether the results obtained in decerebrate fish are relevant for fast processing in behaving rodents. Instead, for awake rodents, we are suggesting that decorrelation in the olfactory bulb is the consequence of the demixing performed by the olfactory bulb–piriform cortex loop, as opposed to a pattern separation mechanism.

Comparison to other models

Other demixing algorithms have been proposed; here we compare to two template-based approaches. One is direct template matching, in which the concentration of an odor is given by the normalized

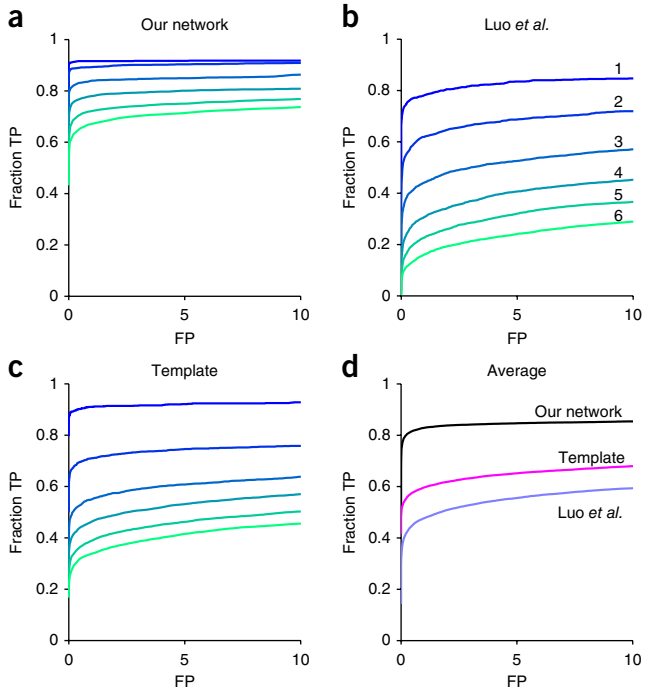

Figure 7 Generalized ROC curves. Curves represent fraction of odors correctly identified (number of true positives (TP) versus number of false positives (FP) as the threshold varies); see text for details. In **a**–**c**, colors indicate number of odors present; the labeling in **b** shows the mapping between color and number. (**a**) Our network. (**b**) Fisher's linear discriminant. (**c**) Template matching. (**d**) Average ROC curve, weighted by the probability that an odor is present, for the three methods.

dot product between the ORN responses and a template—a common approach in the olfactory system[31]. The other is Fisher's linear discriminant, which is a sophisticated version of template matching: ORN responses are passed through a nonlinearity, and then distance to a template is computed using a metric that takes into account the pattern correlations. The latter was used recently in the analysis of a model of the fly olfactory system[32].

We can compare these algorithms to ours using a generalization of receiver operating characteristic (ROC) curves: we plot hit rate (correct identification of odors) versus false positives as the threshold for detecting an odor changes. (Note that this analysis differs from that in **Fig. 5**, where the threshold was fixed.) Because more than one odor can be present at a time, we define the hit rate as the fraction of odors that were correctly identified (see Online Methods section "ROC curves"). Our method does significantly better than the others (**Fig. 7**). This is not surprising, as our method performs approximate Bayesian inference, whereas the other methods are more *ad hoc*.

There are, of course, other approaches to demixing odors on a single sniff cycle. Most of them effectively choose the odor concentrations that best match the observed firing rates, with a regularizing term that prevents overfitting[19,33,34]. Our model falls into this class, with both the definition of "best match" and the regularizing term derived using approximate Bayesian inference (see Online Methods section "Regularized models").

DISCUSSION

Olfaction as probabilistic inference

The olfactory system, like all sensory systems, is faced with an inference problem: given the activity of the olfactory receptor neurons, the brain can never know exactly what odors are present. It can, though, do the next best thing: it can provide a probability distribution over odors. Given that distribution, it can make informed decisions ("Do I eat that piece of fruit that smells like a grapefruit but looks like an orange?"). We thus treated olfaction as a probabilistic inference problem. For that we took the standard Bayesian approach: we wrote down an encoding model—a probabilistic transformation from odors to neural activity—and inverted that transformation using Bayes' theorem. As is typical of realistic problems, the second step could not be done exactly, so we had to compute an approximate posterior distribution over concentration.

The approximate inference algorithm we used matched what is seen in the early olfactory system in several ways. First, it led to circuitry that is broadly consistent with the anatomy of the olfactory bulb: we could naturally identify mitral cells and granule cells, those two cell types were reciprocally connected, the mitral cells projected to piriform cortex, and cells in piriform cortex projected back to granule cells. Second, our simulations produced, for a range of olfactory scenes and conditions, firing patterns that were consistent with *in vivo* activity. Third, our model matched behavioral data in which mice were asked to extract a target odor from up to 13 distractors. We note, though, that for this task Bayesian inference did not appear to be necessary: a linear decoder applied to glomeruli activity also matched the behavioral data[27]. This is at least partially because the concentration of the target and background odors were constant across trials, which simplifies considerably the segmentation problem. Thus, it is hard to distinguish the two models solely on the basis of this experiment. However, the anatomy of the olfactory bulb argues against a linear decoder of glomerular activity, and a linear decoder does not do as well as our model (compare **Fig. 7a** and **c**). Finally, activity patterns of the mitral cells evolved in a way consistent with pattern decorrelation[28]. However, in our network pattern decorrelation is not synonymous with pattern separation, in the sense that it does not mean odors are easier to discriminate over time. Instead, decorrelation is a byproduct of our demixing algorithm.

Previous work

Ours is not, of course, the only model of the olfactory system; many have been proposed, with varying goals[5–9,19,33–47]. Early theoretical work focused on reproducing the oscillations seen in the olfactory bulb in response to single odors, either with[35] or without[37–40] associative memory. Li[5–7] was the first to address the problem of multiple odors, although she considered a situation in which odors are added on each sniff cycle. Hendin *et al.*[41] considered a one-sniff, one-odor version of this model, but with learning; later they extended it to allow the detection of one odor within a mixture[6]. Other models that perform demixing on a single breathing cycle either assume that different odors have independent, and non-Gaussian, temporal fluctuations[8,9] or require a rather elaborate spike timing scheme[42,48]; in both cases, it is unclear whether these assumptions are realistic. Thus, to our knowledge, our model is the first one to identify a cortical–bulbar circuit that computes a probability distribution over odors, and does so in one breathing cycle without assuming independent fluctuations or relying on precise spike timing.

Other models have treated the olfactory bulb as a preprocessing step, one that makes it easier for downstream structures to infer what odor is present[19,33,34,43]. Of these, the closest to our work are a set of models that both sparsify responses and reduce redundancy[19,33,34]. In the hands of Koulakov and Rinberg[19], this was accomplished by reciprocal connections between mitral cells and granule cells. The other two studies[33,34] adopted different implementations, but the underlying equations were very similar (see Online Methods section "Regularized

models"). Alternatively, some models use a form of lateral inhibition, in which mitral cells inhibit each other locally and inhibition is stronger between cells that are neighbors in olfactory space[44–46]. Besides leading to sparse activity, these models tend to decorrelate responses, in the sense that responses to different odors become less similar. Finally, one model uses randomly connected recurrent networks to decorrelate responses[47], indicating that decorrelation does not need special circuitry. While these models result in redundancy reduction, it is not immediately clear how, or even whether, they help with odor segmentation. In particular, to our knowledge these model have not been used to demix a complex odor scene.

Experimental predictions

Our model makes several testable predictions. First, inhibition of mitral cells by granule cells should be divisive (see equation (4)). Testing this prediction experimentally is difficult, as it requires knowledge of connectivity and whole cell *in vivo* recordings. However, it should be feasible in the not so distant future.

Second, our model predicts that when an unknown odor is presented, many granule cells should fire, but at low rates. In contrast, when a known odor is presented, a small fraction of granule cells should fire, but at relatively high rates. This could be assessed using standard methods, such as calcium imaging in awake animals; there is, in fact, already some evidence for it[49]. Our model predicts essentially the same thing in piriform cortex: known odors should activate a relatively small fraction of neurons, but at high firing rates; an unknown odor, on the other hand, should activate a much larger fraction of neurons, but at much lower firing rate. This too could be assessed using calcium imaging.

Li and Hertz[7] also predict different responses for known versus unknown odors. However, in their model those differences do not arise until after the first sniff cycle: the first sniff produces a strong response whether the odor is known, unknown, or a mixture of known and unknown odors; it is only on subsequent sniffs that the response in the bulb is suppressed, and then only if there is a single, known odor. Experimental tests of this prediction, which would go a long way toward distinguishing our model from that of Li and Hertz[7], would be relatively straightforward. (For additional experimental predictions of Li's model, many of which have been verified, see ref. 50.)

Broader implications

The set of difficulties faced by the olfactory system is not unique to that system; it is common to all sensory modalities: we never observe objects in isolation, and the transformation from stimulus to neural response is noisy and often ambiguous. For example, we never observe visual objects against a uniform black background; instead, objects appear against complex backgrounds, and they are often partially occluded. Moreover, we see a two-dimensional representation of the world from which we want to infer three dimensional shapes, a fundamentally ambiguous process. Thus, the lessons we learn from olfaction are likely to provide insight into other sensory modalities.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

Funding for A.G.-B. and P.E.L. was provided by the Gatsby Charitable Foundation; for Z.F.M. and A.P., by the Human Frontiers Science Programme (RGP0027/2010) and the Simons Foundation (325057); for A.P., by the Swiss National Foundation (31003A_143707).

AUTHOR CONTRIBUTIONS

A.G.-B., Z.F.M., A.P. and P.E.L conceived the project. A.G.-B., S.B., J.B., A.P. and P.E.L. developed the theory. A.G.-B., S.B., A.P. and P.E.L. wrote the manuscript. A.G.-B. performed the simulations.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Su, C.-Y., Menuz, K. & Carlson, J.R. Olfactory perception: receptors, cells, and circuits. *Cell* **139**, 45–59 (2009).
2. Rokni, D., Hemmelder, V., Kapoor, V. & Murthy, V.N. An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nat. Neurosci.* **17**, 1225–1232 (2014).
3. Malnic, B., Hirono, J., Sato, T. & Buck, L.B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
4. Pouget, A., Beck, J.M., Ma, W.J. & Latham, P.E. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16**, 1170–1178 (2013).
5. Li, Z. A model of olfactory adaptation and sensitivity enhancement in the olfactory bulb. *Biol. Cybern.* **62**, 349–361 (1990).
6. Hendin, O., Horn, D. & Tsodyks, M.V. Associative memory and segmentation in an oscillatory neural model of the olfactory bulb. *J. Comput. Neurosci.* **5**, 157–169 (1998).
7. Li, Z. & Hertz, J. Odour recognition and segmentation by a model olfactory bulb and cortex. *Network* **11**, 83–102 (2000).
8. Hopfield, J.J. Olfactory computation and object perception. *Proc. Natl. Acad. Sci. USA* **88**, 6462–6466 (1991).
9. Hendin, O., Horn, D. & Hopfield, J.J. Decomposition of a mixture of signals in a model of the olfactory bulb. *Proc. Natl. Acad. Sci. USA* **91**, 5942–5946 (1994).
10. Connelly, T., Savigner, A. & Ma, M. Spontaneous and sensory-evoked activity in mouse olfactory sensory neurons with defined odorant receptors. *J. Neurophysiol.* **110**, 55–62 (2013).
11. Beck, J., Heller, K. & Pouget, A. Complex inference in neural circuits with probabilistic population codes and topic models. in *Advances in Neural Information Processing Systems* **25** (Curran Associates, 2012).
12. Grabska-Barwińska, A., Beck, J., Pouget, A. & Latham, P. Demixing odors – fast inference in olfaction. in *Advances in Neural Information Processing Systems* **26** (Curran Associates, 2013).
13. Egger, V., Svoboda, K. & Mainen, Z.F. Mechanisms of lateral inhibition in the olfactory bulb: efficiency and modulation of spike-evoked calcium influx into granule cells. *J. Neurosci.* **23**, 7551–7558 (2003).
14. Weiss, Y., Simoncelli, E.P. & Adelson, E.H. Motion illusions as optimal percepts. *Nat. Neurosci.* **5**, 598–604 (2002).
15. Shepherd, G. (ed.) *The Synaptic Organization of the Brain* 5th edn. (Oxford Univ. Press, 2004).
16. Fukunaga, I., Berning, M., Kollo, M., Schmaltz, A. & Schaefer, A.T. Two distinct channels of olfactory bulb output. *Neuron* **75**, 320–329 (2012).
17. Igarashi, K.M. *et al.* Parallel mitral and tufted cell pathways route distinct odor information to different targets in the olfactory cortex. *J. Neurosci.* **32**, 7970–7985 (2012).
18. Spors, H. *et al.* Illuminating vertebrate olfactory processing. *J. Neurosci.* **32**, 14102–14108 (2012).
19. Koulakov, A.A. & Rinberg, D. Sparse incomplete representations: a potential role of olfactory granule cells. *Neuron* **72**, 124–136 (2011).
20. Fuentes, R.A., Aguilar, M.I., Aylwin, M.L. & Maldonado, P.E. Neuronal activity of mitral-tufted cells in awake rats during passive and active odorant stimulation. *J. Neurophysiol.* **100**, 422–430 (2008).
21. Cury, K.M. & Uchida, N. Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* **68**, 570–585 (2010).
22. Shusterman, R., Smear, M.C., Koulakov, A.A. & Rinberg, D. Precise olfactory responses tile the sniff cycle. *Nat. Neurosci.* **14**, 1039–1044 (2011).
23. Gschwend, O. *et al.* Neuronal pattern separation in the olfactory bulb improves odor discrimination learning. *Nat. Neurosci.* **18**, 1474–1482 (2015).
24. Moulton, D. Electrical activity in the olfactory system of rabbits with indwelling electrodes. in *Wenner-Gren Center International Symposium Series* vol. 1, 71–84 (Pergamon, 1963).
25. Otazu, G.H., Chae, H., Davis, M.B. & Albeanu, D.F. Cortical feedback decorrelates olfactory bulb output in awake mice. *Neuron* **86**, 1461–1477 (2015).
26. Jinks, A. & Laing, D.G. A limit in the processing of components in odour mixtures. *Perception* **28**, 395–404 (1999).
27. Mathis, A., Rokni, D., Kapoor, V., Bethge, M. & Murthy, V.N. Reading out olfactory receptors: Feedforward circuits detect odors in mixtures without demixing. *Neuron* **91**, 1110–1123 (2016).
28. Friedrich, R.W. & Laurent, G. Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science* **291**, 889–894 (2001).

29. Gschwend, O., Beroud, J. & Carleton, A. Encoding odorant identity by spiking packets of rate-invariant neurons in awake mice. *PLoS One* **7**, e30155 (2012).
30. Uchida, N. & Mainen, Z.F. Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.* **6**, 1224–1229 (2003).
31. Shen, K., Tootoonian, S. & Laurent, G. Encoding of mixtures in a simple olfactory system. *Neuron* **80**, 1246–1262 (2013).
32. Luo, S.X., Axel, R. & Abbott, L.F. Generating sparse and selective third-order responses in the olfactory system of the fly. *Proc. Natl. Acad. Sci. USA* **107**, 10713–10718 (2010).
33. Druckmann, S., Hu, T. & Chklovskii, D.B. A mechanistic model of early sensory processing based on subtracting sparse representations. in *Advances in Neural Information Processing Systems* **25**, 1979–1987 (Curran Associates, 2012).
34. Tootoonian, S. & Lengyel, M. A dual algorithm for olfactory computation in the locust brain. In *Advances in Neural Information Processing Systems* **27**, 2276–2284 (Curran Associates, 2014).
35. Baird, B. Nonlinear dynamics of pattern formation and pattern recognition in rabbit olfactory bulb. *Physica* **22D**, 150–175 (1986).
36. Erdi, P., Gröbler, T., Barna, G. & Kaski, K. Dynamics of the olfactory bulb: bifurcations, learning, and memory. *Biol. Cybern.* **69**, 57–66 (1993).
37. Freeman, W.J. Nonlinear dynamics of paleocortex manifested in the olfactory EEG. *Biol. Cybern.* **35**, 21–37 (1979).
38. Freeman, W.J. EEG analysis gives model of neuronal template-matching mechanism for sensory search with olfactory bulb. *Biol. Cybern.* **35**, 221–234 (1979).
39. Li, Z. & Hopfield, J.J. Modeling the olfactory bulb and its neural oscillatory processings. *Biol. Cybern.* **61**, 379–392 (1989).
40. Yao, Y. & Freeman, W.J. Model of biological pattern recognition with spatially chaotic dynamics. *Neural Netw.* **3**, 153–170 (1990).
41. Hendin, O., Horn, D. & Tsodyks, M.V. The role of inhibition in an associative memory model of the olfactory bulb. *J. Comput. Neurosci.* **4**, 173–182 (1997).
42. Brody, C.D. & Hopfield, J.J. Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron* **37**, 843–852 (2003).
43. Polese, D., Martinelli, E., Marco, S., Di Natale, C. & Gutierrez-Galvez, A. Understanding odor information segregation in the olfactory bulb by means of mitral and tufted cells. *PLoS One* **9**, e109716 (2014).
44. Arevian, A.C., Kapoor, V. & Urban, N.N. Activity-dependent gating of lateral inhibition in the mouse olfactory bulb. *Nat. Neurosci.* **11**, 80–87 (2008).
45. Cleland, T.A. Early transformations in odor representation. *Trends Neurosci.* **33**, 130–139 (2010).
46. Cleland, T.A. & Linster, C. On-center/inhibitory-surround decorrelation via intraglomerular inhibition in the olfactory bulb glomerular layer. *Front. Integr. Neurosci.* **6**, 5 (2012).
47. Wiechert, M.T., Judkewitz, B., Riecke, H. & Friedrich, R.W. Mechanisms of pattern decorrelation by recurrent neuronal circuits. *Nat. Neurosci.* **13**, 1003–1010 (2010).
48. Hopfield, J.J. Odor space and olfactory processing: collective algorithms and neural implementation. *Proc. Natl. Acad. Sci. USA* **96**, 12506–12511 (1999).
49. Kato, H.K., Chu, M.W., Isaacson, J.S. & Komiyama, T. Dynamic sensory representations in the olfactory bulb: modulation by wakefulness and experience. *Neuron* **76**, 962–975 (2012).
50. Zhaoping, L. Olfactory object recognition, segmentation, adaptation, target seeking, and discrimination by the network of the olfactory bulb and cortex: computational model and experimental data. *Curr. Opin. Behav. Sci.* **11**, 30–39 (2016).

ONLINE METHODS

Introduction. Here we provide details of our approximate inference algorithm, which leads ultimately to a set of equations that can be implemented in a neuronal network (equation (28)), with parameters given in section "Parameters." We then explain our method for determining the probability that an odor is present, how we compute ROC curves, and the relationship of our approach to other regularized models.

Exact inference. As in all Bayesian models, inference in olfaction requires that we invert the generative model and write down the probability distribution over concentration given spikes from the olfactory receptor neurons (ORNs). We start by writing down the generative model and discussing exact inference.

The generative model consists of two parts: the likelihood, which is the probability of the data (in our case, spike trains from ORNs) given the odors that are present, and the prior probability distribution over odors. We already have an expression for the prior (equation (2)), so we just need the likelihood. In our model, ORNs generate spikes via a Poisson process. We assume that there is an effective time window, denoted Δt, for counting spikes; thus, the relevant quantity is the probability distribution over spike counts in that time window. Using r_i to denote the spike count of ORNs of receptor type i in time Δt, the distribution over $\mathbf{r}$ ($\equiv r_1, r_2, \ldots, r_N$) is given by

$$p(\mathbf{r}\,|\,\mathbf{c}) = \prod_{i=1}^{N} \frac{1}{r_i!}\left(\sum_{j=0}^{K} w_{ij}c_j\right)^{r_i} \exp\left(-\sum_{j=0}^{K} w_{ij}c_j\right) \tag{6}$$

where N is the number of ORN receptor types and K is the number of odors (see section "Parameters"). To take into account the background activity, the sum over j runs from 0 to K (instead of 1 to K, as it did in equation (1)), and we have defined

$$w_{i0}c_0 \equiv \nu_{0,\,i}\Delta t \tag{7}$$

The likelihood, equation (6), came from the encoding model given in equation (1). In that encoding model, we assumed that ORNs respond linearly to log concentration. While this is a relatively common assumption[2,6,33,36,43], it is an approximation. It is certainly valid over one log unit, as individual ORNs are approximately linear over that range[51]; diversity within an ORN receptor type may extend that to two or three log units[52–54]. A model taking into account nonlinearities would be interesting, but it is beyond the scope of this work.

Combining equation (6) with the expression for the prior, equation (2), the distribution over concentration given spike count is

$$p(\mathbf{c}\,|\,\mathbf{r}) \propto \prod_{i}\left(\sum_{j} w_{ij}c_j\right)^{r_i} \exp\left(-\sum_{j=0} w_{ij}c_j\right) \times \prod_{j=1}\left[(1-p_{\text{prior}})\delta(c_j) + p_{\text{prior}}\frac{e^{-c_j/\beta_{\text{prior}}}}{\beta_{\text{prior}}}\right] \tag{8}$$

While this expression is exact, it contains too much information to be useful to an organism; typically what an organism wants to know is whether or not a particular odor is present, not the full posterior distribution over all possible combinations of odors. In addition, the delta functions, which put point masses at zero concentration, complicate the analysis. For instance, we cannot find the most likely set of odors (the maximum a posteriori solution), as is sometimes done for complicated posteriors, because the delta functions imply that zero concentration has infinite probability density. We assume, then, that the quantity of interest is the marginal distribution over each odor; for odor j, that distribution is given by

$$p(c_j\,|\,\mathbf{r}) = \int \mathrm{d}c_1\mathrm{d}c_2\ldots\mathrm{d}c_{j-1}\mathrm{d}c_{j+1}\ldots\mathrm{d}c_K\, p(\mathbf{c}\,|\,\mathbf{r}) \tag{9}$$

Unfortunately, it is not possible to perform that integral exactly. And even if we could perform the integral over the continuous part of the posterior distribution, because of the delta functions the number of integrals that have to be performed is exponential in the number of odors. We must, therefore, resort to approximate inference.

Approximate inference. There are two parts to our approximate inference algorithm. The first is to replace the prior—which is of the 'spike and slab' form[55], notoriously hard to analyze—with a smoother, approximate prior, denoted $p_{\text{approx}}(\mathbf{c})$. For that we use a product of gamma distributions,

$$p_{\text{approx}}(\mathbf{c}) = \prod_{j=1} \frac{c_j^{\alpha_0-1}e^{-c_j/\beta_0}}{\Gamma(\alpha_0)\beta_0^{\alpha_0}} \tag{10}$$

where $\Gamma(\alpha_0)$ is the standard gamma function. This results in an approximate posterior, denoted $p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r})$,

$$p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r}) \propto \prod_{i}\left(\sum_{j} w_{ij}c_j\right)^{r_i} \exp\left(-\sum_{j=0} w_{ij}c_j\right) \prod_{j=1} c_j^{\alpha_0-1}e^{-c_j/\beta_0} \tag{11}$$

We follow a convention in which probabilities that depend only on $\mathbf{c}$ correspond to priors, whereas those that depend on $\mathbf{c}\,|\,\mathbf{r}$ correspond to posteriors.

We use $\alpha_0 = 1/3$, for reasons discussed below (see the text following equation (54) in section "Regularized models"). To ensure that the mean concentration of each odor under the approximate prior, $\alpha_0\beta_0$, is the same as the mean under the exact prior in equation (2), $p_{\text{prior}}\beta_{\text{prior}}$, we chose β_0 to satisfy $\alpha_0\beta_0 = p_{\text{prior}}\beta_{\text{prior}}$. From section "Parameters," $\beta_{\text{prior}} = 3$ and $p_{\text{prior}} = 3/K$, where K is the number of odors; thus, $\beta_0 = 27/K$.

Although this posterior is simpler than the true one, inference is still intractable, in the sense that even if we replaced $p(\mathbf{c}\,|\,\mathbf{r})$ with $p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r})$ on the right hand side of equation (9), we still could not do the integral. We could do maximum a posteriori inference, but then we would lose any notion of uncertainty. Thus, the second part of our approximation is to use a variational approach: we find a parameterized distribution that is as close as possible to $p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r})$, as assessed by the Kullback-Leibler (KL) divergence[56]. That distribution, denoted $q_{\text{var}}(\mathbf{c}\,|\,\mathbf{r})$, minimizes the KL divergence between $q_{\text{var}}(\mathbf{c}\,|\,\mathbf{r})$ and $p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r})$, which is given by

$$D_{\text{KL}}(q_{\text{var}}(\mathbf{c}\,|\,\mathbf{r})\,||\,p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r})) = \int \mathrm{d}\mathbf{c}\; q_{\text{var}}(\mathbf{c}\,|\,\mathbf{r}) \log \frac{q_{\text{var}}(\mathbf{c}\,|\,\mathbf{r})}{p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r})} \tag{12}$$

It turns out that directly minimizing the KL divergence is also hard, mainly because the first term on the right hand side of equation (11) consists of products of a sum whenever $r_i \geq 2$. We can, though, minimize a bound on the KL divergence. To find the bound, we first use the multinomial theorem to write

$$\left(\sum_{j} w_{ij}c_j\right)^{r_i} = \sum_{\mathbf{N}} \Delta\Big(r_i - \sum_{j=0} N_{ij}\Big) \prod_{j=0} \frac{(w_{ij}c_j)^{N_{ij}}}{N_{ij}!} \tag{13}$$

where the sum over $\mathbf{N}$ is a sum over all sets of non-negative integers N_{ij} such that the N_{ij} add to r_i, a restriction that is enforced by the Kronecker delta,

$$\Delta(n) \equiv \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \tag{14}$$

Inserting equation (13) into (11), we have

$$p_{\text{approx}}(\mathbf{c}\,|\,\mathbf{r}) = \sum_{\mathbf{N}} p_{\text{approx}}(\mathbf{c}, \mathbf{N}\,|\,\mathbf{r}) \tag{15}$$

where

$$p_{\text{approx}}(\mathbf{c}, \mathbf{N}\,|\,\mathbf{r}) \propto \prod_{i}\left[\Delta\Big(r_i - \sum_{j=0} N_{ij}\Big) \prod_{j=0} \frac{(w_{ij}c_j)^{N_{ij}}}{N_{ij}!}\, e^{-w_{ij}c_j}\right] \times \prod_{j=1} c_j^{\alpha_0-1}e^{-c_j/\beta_0} \tag{16}$$

To proceed, we introduce a second variational distribution, $q_{\text{var}}(\mathbf{N} \mid \mathbf{r})$. It is straightforward to show that regardless of what this new variational distribution is, we can bound the Kullback-Leibler divergence via

$$D_{KL}(q_{\text{var}}(\mathbf{c} \mid \mathbf{r}) \,||\, p_{\text{approx}}(\mathbf{c} \mid \mathbf{r})) \le \sum_{\mathbf{N}} \int d\mathbf{c}\; q_{\text{var}}(\mathbf{c} \mid \mathbf{r}) q_{\text{var}}(\mathbf{N} \mid \mathbf{r}) \log \frac{q_{\text{var}}(\mathbf{c} \mid \mathbf{r})\, q_{\text{var}}(\mathbf{N} \mid \mathbf{r})}{p_{\text{approx}}(\mathbf{c}, \mathbf{N} \mid \mathbf{r})} \tag{17}$$

Rather than minimizing the true KL divergence with respect to $q_{\text{var}}(\mathbf{c} \mid \mathbf{r})$, we minimize the right hand side of equation (17), a bound on the true KL divergence, with respect to $q_{\text{var}}(\mathbf{c} \mid \mathbf{r})$ and $q_{\text{var}}(\mathbf{N} \mid \mathbf{r})$. To do that, we differentiate with respect to $q_{\text{var}}(\mathbf{c} \mid \mathbf{r})$ and $q_{\text{var}}(\mathbf{N} \mid \mathbf{r})$ and set the resulting expressions to zero; this gives us

$$\log q_{\text{var}}(\mathbf{c} \mid \mathbf{r}) \sim \sum_{\mathbf{N}} q_{\text{var}}(\mathbf{N} \mid \mathbf{r}) \log p_{\text{approx}}(\mathbf{c}, \mathbf{N} \mid \mathbf{r}) \tag{18a}$$

$$\log q_{\text{var}}(\mathbf{N} \mid \mathbf{r}) \sim \int d\mathbf{c}\; q_{\text{var}}(\mathbf{c} \mid \mathbf{r}) \log p_{\text{approx}}(\mathbf{c}, \mathbf{N} \mid \mathbf{r}) \tag{18b}$$

where here, and in what follows, $\sim$ indicates equality up to an additive constant. To complete these equations, we need an expression for the log of the approximate posterior, $p_{\text{approx}}(\mathbf{c},\mathbf{N} \mid \mathbf{r})$. Using equation (16), that expression is given by

$$\log p_{\text{approx}}(\mathbf{c},\mathbf{N} \mid \mathbf{r}) \sim \sum_i \left[\log \Delta \left(r_i - \sum_{j=0} N_{ij} \right) + \sum_{j=0} N_{ij} \log(w_{ij} c_j) - w_{ij} c_j - \log(N_{ij}!) \right] + \sum_{j=1} (\alpha_0 - 1) \log c_j - c_j / \beta_0 \tag{19}$$

Inserting this into equation (18) yields

$$\log q_{\text{var}}(\mathbf{c} \mid \mathbf{r}) \sim \sum_{j=1} \left[\left(\alpha_0 - 1 + \sum_i \langle N_{ij} \rangle \right) \log c_j - \left(\sum_i w_{ij} + \frac{1}{\beta_0} \right) c_j \right] \tag{20a}$$

$$\log q_{\text{var}}(\mathbf{N} \mid \mathbf{r}) \sim \sum_i \left[\log \Delta \left(r_i - \sum_{j=0} N_{ij} \right) + \sum_{j=0} N_{ij} \langle \log(w_{ij} c_j) \rangle - \log(N_{ij}!) \right] \tag{20b}$$

where the angle brackets indicate an average with respect to either $q_{\text{var}}(\mathbf{c} \mid \mathbf{r})$ or $q_{\text{var}}(\mathbf{N} \mid \mathbf{r})$, whichever is appropriate.

Examining these expressions, we see that $q_{\text{var}}(\mathbf{c} \mid \mathbf{r})$ is gamma distributed and $q_{\text{var}}(\mathbf{N} \mid \mathbf{r})$ is multinomial distributed,

$$q_{\text{var}}(\mathbf{c} \mid \mathbf{r}) = \prod_{j=1} \frac{\bar{c}_j^{\,\bar{c}_j / \beta_j - 1} e^{-c_j / \beta_j}}{\Gamma(\bar{c}_j / \beta_j) \beta_j^{\bar{c}_j / \beta_j}} \tag{21a}$$

$$q_{\text{var}}(\mathbf{N} \mid \mathbf{r}) = \prod_i r_i! \Delta \left(r_i - \sum_{j=0} N_{ij} \right) \prod_{j=0} \frac{p_{ij}^{N_{ij}}}{N_{ij}!} \tag{21b}$$

where

$$\bar{c}_j \equiv \beta_j \alpha_0 + \beta_j \sum_i \langle N_{ij} \rangle \tag{22a}$$

$$\beta_j \equiv \frac{\beta_0}{1 + \beta_0 \sum_i w_{ij}} \tag{22b}$$

$$p_{ij} \equiv \frac{e^{\langle \log(w_{ij} c_j) \rangle}}{\sum_{j=0} e^{\langle \log(w_{ij} c_j) \rangle}} \tag{22c}$$

Note that equation (21a) is the expression that appears in equation (3) of the main text (except that in the main text we do not include the normalization or the subscript "var"). Using equation (21), the averages in equation (22) are readily computed,

$$\langle N_{ij} \rangle = r_i p_{ij} \tag{23a}$$

$$e^{\langle \log(w_{ij} c_j) \rangle} = \begin{cases} \nu_{0,i} \Delta t & j = 0 \\ w_{ij} \beta_j e^{\psi(\bar{c}_j / \beta_j)} & j > 0 \end{cases} \tag{23b}$$

where ψ is the digamma function,

$$\psi(\alpha) \equiv \frac{d \log \Gamma(\alpha)}{d\alpha} \tag{24}$$

Finally, inserting equations (23a), (23b) and (22c) into (22a), we arrive at the update equation for $\bar{c}_j$ (with $j \ge 1$),

$$\bar{c}_j = \beta_j \alpha_0 + \beta_j \sum_i \frac{r_i w_{ij} F_j(\bar{c}_j)}{\nu_{0,i} \Delta t + \sum_k w_{ik} F_k(\bar{c}_k)} \tag{25}$$

where

$$F_j(\bar{c}_j) \equiv \beta_j e^{\psi(\bar{c}_j / \beta_j)} \tag{26}$$

The function $F_j(\bar{c}_j)$ has a relatively simple dependence on $\bar{c}_j$ and β_j. Because $e^{\psi(x)} \approx x^2/2$ if $x < 1$ and $x - 1/2$ if $x \ge 1$, we have

$$F_j(\bar{c}_j) \approx \begin{cases} \bar{c}_j^{\,2} / 2\beta_j & \bar{c}_j < \beta_j \\ \bar{c}_j - \beta_j / 2 & \bar{c}_j \ge \beta_j \end{cases} \tag{27}$$

Equation (25) is appropriate for simulations on a digital computer. However, because we are ultimately interested in a model of the mammalian olfactory system, we need a set of equations that is approximately consistent with known anatomy and physiology. And, of course, those equations must have the same fixed point (or points) as equation (25). Finding such a set of equations is a bit of an art, as they are not unique. The ones we found are

$$\tau_c \frac{d\bar{c}_j}{dt} = \beta_j \alpha_0 - \bar{c}_j + \beta_j F_j(\bar{c}_j) \sum_i \gamma_i^{-1} m_i^2 w_{ij} \tag{28a}$$

$$\tau_m \frac{dm_i}{dt} = -m_i^2 \nu_{0,i} \Delta t + \gamma_i r_i - m_i \sum_k w_{ik}^{mg} g_{ik} \tag{28b}$$

$$\tau_g \frac{dg_{ik}}{dt} = -g_{ik} + g_k w_{ki}^{gm} m_i \tag{28c}$$

$$\tau_g \frac{dg_k}{dt} = -g_k + \sum_j A_{kj} F_j(\bar{c}_j) \tag{28d}$$

where the γ_i are arbitrary positive constants and the weights, $\mathbf{w}^{mg}$, $\mathbf{w}^{gm}$ and $\mathbf{A}$, satisfy

$$w_{ij} = \sum_k w_{ik}^{mg} w_{ki}^{gm} A_{kj} \tag{29}$$

In these equations we interpret g_k as the activity of the soma of granule cell k and g_{ik} as the activity of spine i associated with granule cell k. Dendro-dendritic connections are made at the spines.

To determine the initial conditions, we first ran the model for 2 s with all concentrations set to zero at the beginning of the run, and recorded the values of all dynamical variables ($\bar{c}_j$, m_i, g_{ik} and g_k) at the end of the run. Then, at the

doi:10.1038/nn.4444

start of each simulation, we took the recorded values and added to each of them Gaussian noise with s.d. set to 10% of their value.

To show that in steady state, when all time derivatives are zero, these equations reduce to equation (25), we first use the fact that in steady state, $g_{ik} = g_k w^{\mathrm{gm}}_{ki} m_i$ (equation (28c)). Inserting that into equation (28b) gives, again in steady state,

$$m_i^2 = \frac{\gamma_i r_i}{\nu_{0,i}\Delta t + \sum_k w^{\mathrm{mg}}_{ik} w^{\mathrm{gm}}_{ki} g_k} \tag{30}$$

Thus, in our model granule cells are divisive. Using the steady state solution to equation (28d) to express g_k in terms of the $\bar{c}_j$ and inserting the resulting expression into equation (28a), we arrive at equation (25).

In the main text (section "A probabilistic model of olfaction"), we describe how our network performs inference. The basis for that description is the behavior of the mitral cells, the m_i, and the mean concentration cells, the $\bar{c}_j$. For the mitral cells we need only the steady state behavior; for that we simply insert the steady state solution of equation (28d) into (30), yielding

$$\gamma_i^{-1} m_i^2 = \frac{r_i}{\nu_{0,i}\Delta t + \sum_k w_{ik} F_j(\bar{c}_j)} \tag{31}$$

That corresponds to equation (4), except that in the main text we approximated $F_j(\bar{c}_j)$ by $\bar{c}_j$, an approximation that is valid so long as $\bar{c}_j$ is sufficiently large (see equation (27)). For the mean concentration cells we need the time evolution, for which we can use equation (28a) directly. That correspond to equation (5), except that again we approximated $F_j(\bar{c}_j)$ by $\bar{c}_j$.

Parameters. The parameters we used in our simulations are as follows (see below for a description of the weights):

- $K = 640$, number of odors and number of mean concentration cells ($\bar{c}_j$)
- $N = 160$, number of olfactory receptor neuron (ORN) receptor types and number of mitral cells
- $N_g = 480$, number of granule cells
- $p_{\mathrm{prior}} = 3/K$, prior probability that any particular odor will appear; see equation (2)
- $\beta_{\mathrm{prior}} = 3$, prior over the concentration for present odors; see equation (2)
- $(\alpha_0, \beta_0) = (1/3, 27/K)$, parameters of the gamma distribution used for variational inference; see equation (10)
- $\Delta t = 50$ ms, time window for counting spikes
- $\nu_{0,i}$, background firing rate, drawn from a Gaussian distribution; mean = 10 Hz and s.d. = 1 Hz
- γ_i, variability parameter, drawn from a log normal distribution; mean of $\log\gamma = 0.5$ and s.d. = 0.275
- $\tau_c = 10$ ms, time constant of mean concentration cells
- $\tau_m = 10$ ms, time constant of mitral cells
- $\tau_g = 5$ ms, time constant of granule cells
- $w_{\mathrm{mg}} = 1/\sqrt{20}$, mitral ↔ granule connection strength
- $p_{\mathrm{mg}} = 1/2$, granule cell ↔ mitral cell connection probability
- $A_0 = 15$, cortex → granule cell connection strength
- $p_A = 0.2$, cortex → granule cell connection probability
- $\mathrm{d}t = 0.1$ ms, time step used in simulations

There are three sets of weights we need to specify: w^{mg}_{ik} (granule cell → mitral cell), w^{gm}_{ki} (mitral cell → granule cell), A_{ik} (piriform cortex → granule cell). A diagram showing a qualitative picture of the weights is given in **Figure 1**. In words: there are three main granule cells associated with each mitral cell; for those the probability of a reciprocal connection is 1. In addition, there are three secondary granule cells on either side of the three main granule cells; for those the probability of a reciprocal connection to the same mitral cell is 1/2. Feedback from the mean concentration cells is sparse: the connection probability to granule cells is 0.2, but if there is a connection, it goes to all three main granule cells. A quantitative, but harder to interpret, description follows.

Granule cell → mitral cell:

$$w^{\mathrm{mg}}_{ik} = w_{\mathrm{mg}} \begin{cases} 1 & \left|3i - k \bmod N_g\right| \le 1 \\ \xi^{\mathrm{mg}}_{ik} & 2 \le \left|3i - k \bmod N_g\right| \le 4 \\ 0 & 5 \le \left|3i - k \bmod N_g\right| \end{cases} \tag{32}$$

where

$$\xi^{\mathrm{mg}}_{ik} = \begin{cases} 1 & \text{probability } p_{\mathrm{mg}} \\ 0 & \text{probability } 1 - p_{\mathrm{mg}} \end{cases} \tag{33}$$

Mitral cell → granule cell:

$$w^{\mathrm{gm}}_{ki} = w^{\mathrm{mg}}_{ik} \tag{34}$$

Piriform cortex → granule cell:

$$A_{kj} = A_0 \xi^A_{\lfloor (k+1)/3 \rfloor, j} \tag{35}$$

where $\lfloor \cdot \rfloor$ indicates the integer part and

$$\xi^A_{j'j} = \begin{cases} 1 & \text{probability } p_A \\ 0 & \text{probability } 1 - p_A \end{cases} \tag{36}$$

In the above, i and j range from 1 to N and k from 1 to N_g, both inclusive.

In section "Regularized models" we need the typical value of β_j, so here we compute an approximation to its mean. Given the above definitions, it is straightforward to show that

$$\left\langle \sum_{i=1}^{N} w_{ij} \right\rangle = 3N w^2_{\mathrm{mg}} (1 + 2p_{\mathrm{mg}}) A_0 p_A = 144 \tag{37}$$

where the numerical value of 144 follows by using the parameters given at the beginning of this section. Combining this with the definition of β_j (equation (22b)), and noting that $\beta_0 = 27/K$, we see that

$$\langle \beta_j \rangle \approx \frac{1}{640/27 + 144} \approx \frac{1}{168} \tag{38}$$

Determining the probability that an odor is present. Our inference algorithm takes ORN input and returns a set of $\bar{c}_j$, one for each odor. The olfactory system must then turn each $\bar{c}_j$ into the probability that odor j is present. This probability can be computed from Bayes' theorem,

$$p(1_j \mid \bar{c}_j) = \frac{p(\bar{c}_j \mid 1_j) p(1_j)}{p(\bar{c}_j \mid 1_j) p(1_j) + p(\bar{c}_j \mid 0_j) p(0_j)} \tag{39}$$

where 1_j indicates that odor j is present and 0_j indicates that it is absent. The prior probability that an odor is present, $p(1_j)$, is just p_{prior}, independent of j (see equation (2)). Thus, we need only compute $p(\bar{c}_j \mid 1_j)$ and $p(\bar{c}_j \mid 0_j)$. To do that, we ran 10,000 simulations of equation (28) with the number of presented odors and their concentrations drawn from the prior, equation (2). The $\bar{c}_j$ were evaluated in steady state (at $t = 300$ ms) and binned in log space (we used log space because the $\bar{c}_j$ can be very small), and the probability distributions were estimated from the histograms. Specifically, letting

$$\bar{c}_j^{(k)} = \text{value of } \bar{c}_j \text{ on simulation } k \tag{40a}$$

$$s_j^{(k)} = 1 \text{ if odor } j \text{ was present on simulation } k \text{ and 0 otherwise} \tag{40b}$$

an estimate of the conditional probability distributions over $\log \bar{c}_j$ is found from

$$p(\log \bar{c} \,|\, 1) = \frac{\sum_{jk} s_j^{(k)} \mathrm{I}\left[\log \bar{c} - \delta \log \bar{c}/2 < \log \bar{c}_j^{(k)} \le \log \bar{c} + \delta \log \bar{c}/2\right]}{\sum_{jk} s_j^{(k)}} \tag{41a}$$

$$p(\log \bar{c} \,|\, 0) = \frac{\sum_{jk} (1 - s_j^{(k)}) \mathrm{I}\left[\log \bar{c} - \delta \log \bar{c}/2 < \log \bar{c}_j^{(k)} \le \log \bar{c} + \delta \log \bar{c}/2\right]}{\sum_{jk} (1 - s_j^{(k)})} \tag{41b}$$

where I (·) is the indicator function: I [x] = 1 if x is true and 0 otherwise. The bin size, $\delta \log \bar{c}$, was 0.0182. Because $p(\log \bar{c}) = \bar{c}\, p(\bar{c})$, we can use the conditional distributions over $\log \bar{c}$ to compute $p(1 \,|\, \bar{c})$, the probability that odor 1 is present given $\bar{c}_j$:

$$p(1 \,|\, \bar{c}) = \frac{p_{\text{prior}}\, p(\log \bar{c} \,|\, 1) / p(\log \bar{c} \,|\, 0)}{1 - p_{\text{prior}} + p_{\text{prior}}\, p(\log \bar{c} \,|\, 1) / p(\log \bar{c} \,|\, 0)} \tag{42}$$

We used this expression to compute the points in **Figure 3a**. We then fit those points to a sigmoidal function of the form

$$p(1 \,|\, \bar{c}) = \frac{e^{\kappa(\log \bar{c} - \log \bar{c}_0)}}{1 + e^{\kappa(\log \bar{c} - \log \bar{c}_0)}} \tag{43}$$

Minimizing the mean square error between the data and the curve gave $\kappa = 8.3$ and $\bar{c}_0 = 0.31$; equation (43), with κ and $\bar{c}_0$ set to these values, is the solid green line in **Figure 3a**.

We used the same data to determine the degree to which the model exhibits concentration invariance: we binned the true concentration, and in each bin computed the mean and s.d. of the inferred mean concentration, $\bar{c}_j$. The results are plotted in **Figure 3b**,**c**. The bins were 0.5 in **Figure 3b** and 0.1 in **Figure 3c**.

ROC curves. We constructed three sets of ROC curves, from three different models: one from our simulations (**Fig. 7a**), one from a model by Luo *et al.*[30] (**Fig. 7b**), and one from template matching (**Fig. 7c**). In all three cases, the ROC curves were constructed as follows. The ORN input was mapped to a set of variables, which for now we will call z_j, $j = 1, \ldots, K$ (see below for explicit examples). Odor j was then declared to be present if z_j was above a threshold and absent if it was below the threshold. For each value of the threshold, we computed the fraction of odors that were declared to be present and actually were present (fraction of true positives, TP), and the number of odors that were declared to be present but were not (number of false positives, FP). These two quantities, fraction of true positives and number of false positives, are plotted on the y and x axes, respectively, in **Figure 7**.

What differed among the three models was the mapping from ORN activity to z_j. Our model was simplest: $z_j = \bar{c}_j$ evaluated at the end of the trial (300 ms). For Luo *et al.*[30], the ORN activity on any particular trial was first passed through a nonlinearity to create a new variable, ρ_i,

$$\rho_i = \frac{r_i^n}{r_i^n + \sigma^n + \left((m/K)\sum_{i'} r_{i'}\right)^n} \tag{44}$$

where $n = 1.5$, $\sigma = 2$ and $m = 0.3$ (chosen to maximize performance). Then, defining

$$\boldsymbol{\rho} \equiv (\rho_1, \rho_2, \ldots, \rho_N) \tag{45}$$

(recall that N is the number of ORN receptor types; see section "Parameters"), z_j was given, on any particular trial, by

$$z_j = \frac{(\boldsymbol{\rho}^j - \bar{\boldsymbol{\rho}}) \cdot \mathbf{C}^{-1} \cdot \boldsymbol{\rho}}{(\boldsymbol{\rho}^j - \bar{\boldsymbol{\rho}}) \cdot \mathbf{C}^{-1} \cdot \boldsymbol{\rho}^j} \tag{46}$$

where "·" indicates the standard dot product, ρ^j is the mean activity associated with odor j,

$$\rho_i^j = \int \mathrm{d}c_j \frac{e^{-c_j/\beta_{\text{prior}}}}{\beta_{\text{prior}}} \frac{(w_{ij}c_j)^n}{(w_{ij}c_j)^n + \sigma^n + \left((m/K)\sum_{i'} w_{i'j}c_j\right)^n} \tag{47}$$

and $\bar{\boldsymbol{\rho}}$ and $\mathbf{C}$ are the mean and covariance of $\boldsymbol{\rho}$,

$$\bar{\boldsymbol{\rho}} = \frac{1}{K}\sum_{j=1}^{K} \boldsymbol{\rho}^j \tag{48a}$$

$$\mathbf{C} = \frac{1}{K}\sum_{j=1}^{K} (\boldsymbol{\rho}^j - \bar{\boldsymbol{\rho}})(\boldsymbol{\rho}^j - \bar{\boldsymbol{\rho}}) \tag{48b}$$

For template matching, z_j was taken to be the cosine of the angle between the vector of ORN responses, $\mathbf{r} \equiv (r_1, r_2, \ldots, r_N)$, and the vector of weights $\mathbf{w}_j \equiv (w_{1j}, w_{2j}, \ldots, w_{Nj})$,

$$z_j = \frac{\mathbf{r} \cdot \mathbf{w}_j}{[(\mathbf{r} \cdot \mathbf{r})(\mathbf{w}_j \cdot \mathbf{w}_j)]^{1/2}} \tag{49}$$

For activity, $\mathbf{r}$, we used the vector of spike counts evaluated 50 ms after odor onset.

Regularized models. Our model can be written

$$\bar{\mathbf{c}} = \underset{\bar{\mathbf{c}}'}{\text{argmax}}[\mathrm{L}(\bar{\mathbf{c}}') + \Omega(\bar{\mathbf{c}}')] \tag{50}$$

where

$$\mathrm{L}(\bar{\mathbf{c}}) = \sum_i \left[r_i \log\left(\nu_{0,i}\Delta t + \sum_j w_{ij} F_j(\bar{c}_j)\right) - \sum_j w_{ij} F_j(\bar{c}_j)\right] \tag{51a}$$

$$\Omega(\bar{\mathbf{c}}) = -\sum_j \left[\psi(\bar{c}_j/\beta_j)(\bar{c}_j/\beta_j - \alpha_0) - \log\Gamma(\bar{c}_j/\beta_j) - \sum_i w_{ij} F_j(\bar{c}_j)\right] \tag{51b}$$

To show this, differentiate $\mathrm{L}(\bar{\mathbf{c}}) + \Omega(\bar{\mathbf{c}})$ with respect to $\bar{c}_j$ and set the resulting expression to zero; that yields equation (25). The first term, $\mathrm{L}(\bar{\mathbf{c}})$, is the log likelihood (the log of equation (6)), but with c_j replaced by $F_j(\bar{c}_j)$ (equation (26)), which is the geometric mean under the gamma distribution,

$$F_j(\bar{c}_j) = e^{\langle \log c_j \rangle} \tag{52}$$

The second term, $\Omega(\bar{\mathbf{c}})$, is a regularizer, in the sense that when α_0 is small, it pushes $\bar{c}_j/\beta_j$ toward α_0. To understand how it behaves in that limit, note that when $\bar{c}_j/\beta_j \to 0$,

$$\Gamma(\bar{c}/\beta_j) \to \frac{\beta_j}{\bar{c}_j} \tag{53a}$$

$$\psi(\bar{c}/\beta_j) \to -\frac{\beta_j}{\bar{c}_j} \tag{53b}$$

doi:10.1038/nn.4444

The second expression, combined with the definition of $F_j(\bar{c}_j)$ given in equation (26), implies that $F_j(\bar{c}_j) \to 0$ as $\bar{c}_j / \beta_j \to 0$. Consequently, when both α_0 and $\bar{c}_j / \beta_j$ are small,

$$\Omega(\alpha) \to \sum_j \left[\frac{\bar{c}_j / \beta_j - \alpha_0}{\bar{c}_j / \beta_j} - \log(\bar{c}_j / \beta_j) \right] \tag{54}$$

Each term in brackets has a quadratic peak, with the jth term peaking at $\bar{c}_j = \beta_j$. The fact that the peaks are quadratic makes this different from an L1 regularizer, which would typically drive c_j all the way to 0. Nevertheless, this regularizer pushes c_j close to (but slightly above) $\alpha_0\beta_j$, which is close to zero (about 0.002). The peaks have a width of $\alpha_0\beta_j$; given that $\alpha_0 = 1/3$ (see section "Parameters") and β_j is about 1/168 (see equation (38)), the peaks are relatively narrow and can correspond to sharp local maxima. We chose $\alpha_0 = 1/3$ to avoid getting stuck in these maxima.

Both Koulakov and Rinberg[19] and Tootoonian and Lengyel[32] also cast olfaction as a minimization problem. For both of their models, the likelihood was quadratic,

$$\mathrm{L}(\mathbf{x}) = \frac{1}{2\delta} \sum_i \left(r_i - \sum_j w_{ij} x_j \right)^2 \tag{55}$$

and the regularizer had the form

$$\Omega(\mathbf{x}) = \sum_j \theta_j x_j \Theta(x_j) \tag{56}$$

The variables, x_j, had different interpretations in the two models: for Koulakov and Rinberg it was granule cell activity, and for Tootoonian and Lengyel it was a binary variable signaling the presence or absence of chemicals. In addition, for Tootoonian and Lengyel's model, $\varepsilon \to 0$ (to enforce the constraint exactly), whereas for Koulakov and Rinberg's model, ε was finite.

Druckmann *et al.*[31] considered a very similar model, except they applied it to arbitrary stimuli rather than specifically to olfaction. Their model used the same likelihood as in equation (55), with, as in Tootoonian and Lengyel's, $\varepsilon \to 0$. The regularizer was a combination of L1 and L2 norms,

$$\Omega(\mathbf{x}) = \sum_j |x_j| + \frac{1}{2\delta} \sum_i x_i^2 \tag{57}$$

Unlike in the work targeted specifically for olfaction[19,32], in the model of Druckmann *et al.* the x_j values were not constrained to be positive.

Statistical tests. We performed a statistical test only for **Figure 6b**. The sample size was 36, the number of unique correlation coefficients associated with our 9 randomly chosen odors. The sample size was chosen because it was sufficient to make our point, and is on the same order as the numbers typically used by others in the field. The distribution of the correlations coefficients was assumed to be normal, but this was not formally tested. For other figures, sample size was limited by the number of simulations we were able to run; because this was not a major practical restriction, we ran a large number of simulations, as in other modeling studies.

A **Supplementary Methods Checklist** is available.

Code and data availability. All experimental procedures were based on a custom software, written in the Matlab (R2013a) environment. The code is available as a zipped archive at https://github.com/agnigb/olfaction/. The file README.md contains instructions on how to run the code.

The code uses a different random number seed for each run, so the results will differ slightly from the figures in the paper. This could have been avoided by fixing the random number generator. However, we preferred not to risk drawing conclusions that would not generalize over random seeds.

51. Firestein, S. & Werblin, F. Odor-induced membrane currents in vertebrate-olfactory receptor neurons. *Science* **244**, 79–82 (1989).
52. Wachowiak, M. & Cohen, L.B. Representation of odorants by receptor neuron input to the mouse olfactory bulb. *Neuron* **32**, 723–735 (2001).
53. Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* **22**, 3033–3043 (2002).
54. Grosmaitre, X., Vassalli, A., Mombaerts, P., Shepherd, G.M. & Ma, M. Odorant responses of olfactory sensory neurons expressing the odorant receptor MOR23: a patch clamp analysis in gene-targeted mice. *Proc. Natl. Acad. Sci. USA* **103**, 1970–1975 (2006).
55. Mitchell, T. & Beauchamp, J. Bayesian variable selection in linear regression. *J. Am. Stat. Assoc.* **83**, 1023–1032 (1988).
56. Wainwright, M.J. & Jordan, M.I. *Graphical Models, Exponential Families, and Variational Inference*, Foundations and Trends Machine Learning (Now Publishers, 2008).