Receptor arrays optimized for natural odor statistics

David Zwicker[a,b], Arvind Murugan[a,b,c,d,1], and Michael P. Brenner[a,b,1]

[a]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; [b]Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA 02138; [c]Department of Physics, University of Chicago, Chicago, IL 60637; and [d]James Franck Institute, University of Chicago, Chicago, IL 60637

Edited by Leslie B. Vosshall, The Rockefeller University, New York, NY, and approved March 22, 2016 (received for review January 11, 2016)

Natural odors typically consist of many molecules at different concentrations. It is unclear how the numerous odorant molecules and their possible mixtures are discriminated by relatively few olfactory receptors. Using an information theoretic model, we show that a receptor array is optimal for this task if it achieves two possibly conflicting goals: (*i*) Each receptor should respond to half of all odors and (*ii*) the response of different receptors should be uncorrelated when averaged over odors presented with natural statistics. We use these design principles to predict statistics of the affinities between receptors and odorant molecules for a broad class of odor statistics. We also show that optimal receptor arrays can be tuned to either resolve concentrations well or distinguish mixtures reliably. Finally, we use our results to predict properties of experimentally measured receptor arrays. Our work can thus be used to better understand natural olfaction, and it also suggests ways to improve artificial sensor arrays.

olfaction | sensing | natural statistics | information theory | molecular recognition

Discrimination of olfactory signals occurs in a high-dimensional space of odor stimuli in which a large number of distinct molecules and their mixtures can be distinguished by a much smaller number of receptors (1–3). For example, humans have about 300 distinct olfactory receptors (4), which can sense at least 2,100 odorant molecules (5), and the real number might be much larger (1). Moreover, humans can differentiate between mixtures of up to 30 odorants (6). Such remarkable molecular discrimination is thought to use a combinatorial code (7, 8), where typical odorant molecules bind to receptors of multiple types (1, 3). Each receptor type is expressed in many cells (9), and the information from all receptors of the same type is accumulated in corresponding glomeruli in the olfactory bulb (10, 11) (see Fig. 1*A*). The activity of a single glomerulus is thus the total signal of the associated receptor type, so the information about the odor is encoded in the activity pattern of the glomeruli (11, 12). This activity pattern is interpreted by the brain to learn about the composition and the concentration of the inhaled odor. We here study how receptor arrays can maximize the transmitted information.

It is known (13, 14) that the input–output characteristics of sensory apparatuses of many organisms are tailored to the statistics of the organism's natural environment to maximize information transmission. For example, in the visual circuit of the fly, the input–output relationship of neurons is matched to the cumulative distribution of the input distribution (13). Similar observations have since been made in many sensory systems (14, 15) and even in transcriptional regulation (16). In all these cases, the distinguishable outputs of the sensory system must be dedicated to equal parts of the input distribution, which is known as Laughlin's principle (13) or histogram equalization (17). Intuitively, more of the response range is dedicated to common stimuli, at the expense of less frequent stimuli (13).

Similarly, the binding affinities of olfactory receptors might reflect the natural statistics of odors in an organism's environment. Odors vary across environments and differ in both their frequency and composition (18). For example, some molecules might frequently appear together because they originate from the same source, whereas others are rarely found in the same odor. Additionally, some odors are more important to recognize than others, which corresponds to considering an increased frequency for these odors. Together, the frequencies and correlations constitute the natural olfactory scene.

It is not clear how olfactory receptors can account for natural odor statistics. Merely dedicating more receptors to common odors is not optimal, given the small number of available receptors and the many-to-many relationship between receptors and odors. Further, the value of a receptor is strongly dependent on how it complements the other receptors in the array; many "good" receptors can still create a poor array. Finally, the concentrations of molecules composing an odor can vary widely. Odors need to be distinguished in both quality and quantity; hence receptors must vary in both what molecules they respond to and how strongly they do this. Given the statistics of an olfactory scene, what combination of odorants should different receptors in an array respond to?

We use an information theoretic approach to quantify how well a receptor array is matched to given odor statistics. We generalize Laughlin's principle to the high-dimensional case and show that optimal receptor arrays should obey two general principles: (*i*) Each receptor should be active half the time when odors are presented with natural statistics. (*ii*) The activities of any pair of receptors should be uncorrelated when averaged over all odors presented with natural statistics. If both conditions are satisfied for an array of N_r receptors with binary readouts, all 2^{N_r} activity patterns are equally likely when odors are presented with natural statistics (see Fig. 1*B*). The two basic principles may be obvious with some thought, but they usually cannot be satisfied simultaneously. We thus also determine the relative costs of violating the two conditions and use this to carry out numerical and analytical optimizations to determine conditions for optimal receptor arrays. Furthermore, our model implies relationships between the typical ligand concentrations and the ability to discriminate mixtures that have been missed before.

Significance

Natural odors typically consist of many molecules at different concentrations, which together determine the odor identity. This information is collectively encoded by olfactory receptors and then forwarded to the brain. However, it is unclear how the receptor activity can encode both the composition of the odor and the concentrations of its constituents. We study a simple model of the olfactory receptors from which we derive design principles for optimally communicating odor information in a given natural environment. We use these results to discuss biological olfactory systems, and we propose how they can be used to improve artificial sensor arrays.

Author contributions: D.Z., A.M., and M.P.B. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: brenner@seas.harvard.edu or amurugan@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600357113/-/DCSupplemental.

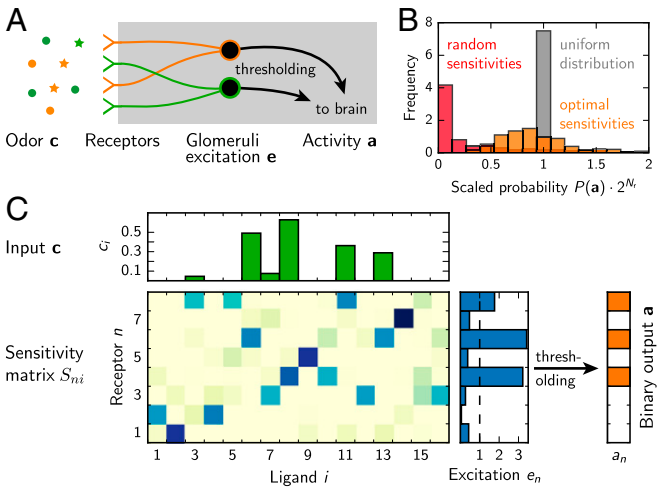

Fig. 1. (*A*) Schematic representation of the olfactory system, where ligands bind to receptors, whose excitation is accumulated in glomeruli, thresholded, and relayed to the brain. (*B*) Histogram of the probabilities $P(\boldsymbol{a})$ of the 2^{N_r} output patterns $\boldsymbol{a}$ for a random receptor array (red, entropy $I = 6.15$ bits), a numerically optimized one (orange, $I = 7.83$ bits), and the theoretical optimum of a uniform distribution (gray, $I = 8$ bits). (*C*) Schematic representation of our physical model, where the input $\boldsymbol{c}$ (green bars) is mapped to excitations (blue bars), which are turned into the output $\boldsymbol{a}$ (orange) by thresholding. Parameters in *B* and *C* are: receptor count $N_r = 8$, ligand count $N_l = 16$, ligand frequency $p_i = 1/4$, and mean $\mu_i = 1$ and SD $\sigma_i = 1$ of the ligand concentration.

After introducing our general framework below, we first discuss general properties of optimal receptor arrays. We then consider two different classes of natural statistics, for which we find optimal receptors in terms of random matrices. Here, our information theoretic approach provides a combined measure of the array's performance in multiple aspects—from the resolution of ligand concentrations to the discrimination of mixture composition. We thus finally discuss the trade-off between such potentially mutually exclusive goals and compare our results to experimentally measured receptor arrays.

Results

Odors are mixtures of odorant molecules that are ligands of olfactory receptors. Any odor can be described by a vector $\boldsymbol{c} = (c_1, c_2, \ldots, c_{N_l})$ that specifies the concentrations c_i of all N_l possible ligands ($c_i \geq 0$). During a single sniff, the ligands in the odor $\boldsymbol{c}$ come in contact with N_r different odor receptors. In the simplest case, the sensitivity of receptor n to ligand i can be described by a single number S_{ni}, and the total excitation e_n of receptor n is given by (19, 20)

$$e_n = \sum_i S_{ni} c_i. \qquad [1]$$

Typical receptors have a nonlinear dose–response curve (21), and the output a_n is thus a nonlinear function of e_n. Moreover, receptors are subject to noise (22), e.g., from stochastic binding, which limits the number of distinguishable outputs. To capture both effects, we consider receptors with only two output states, which corresponds to large noise (23). In this case, the activity a_n of receptor n is given by

$$a_n = \begin{cases} 0 & e_n < 1 \\ 1 & e_n \geq 1 \end{cases}, \qquad [2]$$

i.e., the receptor is active if its excitation e_n exceeds a threshold. Eqs. **1** and **2** describe the mapping of the odor $\boldsymbol{c}$ to the activity pattern $\boldsymbol{a} = (a_1, a_2, \ldots, a_{N_r})$, where the receptor array is characterized by the sensitivity matrix S_{ni} (see Fig. 1*C*). This activity pattern is then analyzed by the brain to infer the odor $\boldsymbol{c}$. Such a distributed representation of odors in activity patterns has been compared with compressed sensing (24); here we focus on how this representation can be tuned to match the structure of natural odors.

We assume that the structure of natural odors in a given environment can be captured by a probability distribution $P_{\text{env}}(\boldsymbol{c})$ from which odors are drawn. $P_{\text{env}}(\boldsymbol{c})$ can encode, for example, the fact that some ligands are more common than others or that some ligands are strongly correlated or anticorrelated in their occurrence. Because natural odor statistics are hard to measure (18), we work with a broad class of distributions $P_{\text{env}}(\boldsymbol{c})$ characterized by a few parameters. We define p_i to be the probability with which ligand i occurs in a random odor. The correlations between the occurrences of ligands are captured by a covariance matrix p_{ij}. We expect p_i to be small because any given natural odor typically contain tens to hundreds of ligands (20, 25), which is a small subset of all $N_l \gtrsim 2{,}100$ ligands (18). When a ligand i is present, we assume its concentration c_i has mean μ_i and standard deviation (SD) σ_i. Thus, the full natural odor statistics $P_{\text{env}}(\boldsymbol{c})$ are parameterized by p_i, μ_i, and σ_i for all ligands i and a covariance matrix p_{ij} in our model.

Optimal Receptor Arrays. An optimal receptor array must tailor receptor sensitivities S_{ni} so that the odors-to-activity mapping given by Eqs. **1** and **2** dedicates more activity patterns to more frequent or more important odors as specified by $P_{\text{env}}(\boldsymbol{c})$. In information theoretic terms, the array must maximize the mutual information $I(\boldsymbol{c}, \boldsymbol{a})$ (26). In our model, the mapping from $\boldsymbol{c}$ to $\boldsymbol{a}$ is deterministic, and I can be written as the entropy of the output distribution $P(\boldsymbol{a})$,

$$I = -\sum_{\boldsymbol{a}} P(\boldsymbol{a}) \log_2 P(\boldsymbol{a}), \qquad [3]$$

where the sum is over all possible activity patterns $\boldsymbol{a}$. Note that $P(\boldsymbol{a}) = \int \mathrm{d}\boldsymbol{c}\, P(\boldsymbol{a}|\boldsymbol{c}) P_{\text{env}}(\boldsymbol{c})$, where $P(\boldsymbol{a}|\boldsymbol{c})$ describes the mapping from $\boldsymbol{c}$ to $\boldsymbol{a}$. Consequently, I depends on S_{ni} and the odor environment $P_{\text{env}}(\boldsymbol{c})$. In fact, I is maximized by sensitivities S_{ni} that are tailored to $P_{\text{env}}(\boldsymbol{c})$ such that all activity patterns $\boldsymbol{a}$ are equally likely (13, 26).

The mutual information I can be approximated (27) in terms of the mean activities $\langle a_n \rangle$ and the covariance between receptors, $\mathrm{cov}(a_n, a_m) = \langle a_n a_m \rangle - \langle a_n \rangle \langle a_m \rangle$, encoded by $P(\boldsymbol{a})$,

$$I \approx -\sum_n [\langle a_n \rangle \log_2 \langle a_n \rangle + (1 - \langle a_n \rangle) \log_2 (1 - \langle a_n \rangle)] - \frac{8}{\ln 2} \sum_{n<m} \mathrm{cov}(a_n, a_m)^2, \qquad [4]$$

which is an expansion up to quadratic order in $\mathrm{cov}(a_n, a_m)$. The first term gives the information gained through each receptor in isolation. The second term describes the reduction of information due to correlations between different receptors. For both Eqs. **3** and **4**, the maximal mutual information of N_r bits can only be obtained if

$$\langle a_n \rangle^* = \frac{1}{2} \qquad [5a]$$

and

$$\mathrm{cov}(a_n, a_m)^* = 0. \qquad [5b]$$

Consequently, in a receptor array optimized for its natural environment, each receptor responds to about half of all odors and any pair of receptors is uncorrelated in its response to odors, assuming odors are presented with frequency $P_{\text{env}}(\boldsymbol{c})$.

These design principles follow from very general considerations, but they may not always be simultaneously achievable. To understand such constraints, we study how microscopic properties of receptor arrays (the sensitivities S_{ni}) determine both $\langle a_n \rangle$ and $\mathrm{cov}(a_n, a_m)$. The mean receptor activity $\langle a_n \rangle$ is given by the probability that the associated excitation e_n exceeds 1, $\langle a_n \rangle = 1 - F_n(1)$, where $F_n(e_n)$ denotes the cumulative distribution function of e_n (see *Supporting Information*). The covariance $\mathrm{cov}(a_n, a_m)$ can be estimated in terms of $\mathrm{cov}_c(e_n, e_m)$ using a normal approximation around the maximum of I (see *Supporting Information*). These statistics of e_n can be calculated from Eq. **1** and read

$$\langle e_n \rangle_c = \sum_i S_{ni} \langle c_i \rangle \qquad \textbf{[6a]}$$

$$\mathrm{cov}_c(e_n, e_m) = \sum_{i,j} S_{ni} S_{mj} \mathrm{cov}(c_i, c_j), \qquad \textbf{[6b]}$$

where $\langle c_i \rangle$ and $\mathrm{cov}(c_i, c_j)$ follow from $P_{\mathrm{env}}(\boldsymbol{c})$.

Combining Eqs. **4** and **6** to estimate mutual information, we can quantify how well an array's sensitivities S_{ni} are matched to natural odor statistics $P_{\mathrm{env}}(\boldsymbol{c})$. As a computational matter, these equations also allow a rapid calculation of mutual information without calculating the full distribution $P(\boldsymbol{a})$.

Random Sensitivity Matrices. We next study which sensitivity matrices S_{ni} obey the optimization goals given in Eq. **5** for given odor statistics. Here, we will show that random S_{ni} with independent and identically distributed entries drawn from the right distribution can be close to optimal. This is because such matrices generically have low correlations, and the resulting activities a_n are thus only weakly correlated. In this section, we study what distributions lead to $\langle a_n \rangle = \frac{1}{2}$ and under what conditions these matrices minimize $\mathrm{cov}(a_n, a_m)$ for two different classes of odor distributions.

Narrow concentration distributions. We begin with the simple case where the concentration distributions are narrow, $\sigma_i \ll \mu_i$. In this case, we can focus on determining which ligands appear in a mixture. Receptors that are optimal for this task must be highly sensitive to some ligands while they ignore the others, but the exact value of the sensitivity does not matter. This property can be encoded in a binary sensitivity matrix $\hat{S}_{ni}$ where $\hat{S}_{ni} = 1$ if receptor n reacts to ligand i and $\hat{S}_{ni} = 0$ if it does not. We can then calculate activity statistics using Eqs. **2** and **6**, as shown in *Supporting Information*. In the simple case of uncorrelated mixtures ($p_{ij} = 0$ for $i \neq j$), $\langle a_n \rangle \approx \sum_i \hat{S}_{ni} p_i$ and $\mathrm{cov}(a_n, a_m) \approx \sum_i \hat{S}_{ni} \hat{S}_{mi} p_i$. In *Supporting Information*, we also calculate corrections due to the correlated appearance of ligands ($p_{ij} \neq 0$); e.g., $\langle a_n \rangle \approx \langle a_n \rangle_0 + \frac{1}{2}(1 - \langle a_n \rangle_0) \sum_{i,j} (\hat{S}_{ni} + \hat{S}_{nj} - \hat{S}_{ni}\hat{S}_{nj}) p_{ij}$, where $\langle a_n \rangle_0 = \sum_i \hat{S}_{ni} p_i$ is the receptor activity in the uncorrelated case.

In the case of uncorrelated mixtures, we find, using Eq. **5**, that $\hat{S}_{ni}$ for optimal receptor arrays must satisfy

$$\sum_i \hat{S}^*_{ni} p_i = \frac{1}{2} \qquad \textbf{[7a]}$$

and

$$\sum_i \hat{S}^*_{ni} \hat{S}^*_{mi} p_i = 0. \qquad \textbf{[7b]}$$

Receptors are thus optimal if (i) the occurrence probabilities p_i of the ligands they react to add up to 1/2 and (ii) no ligand activates multiple receptors. Because any given ligand is rare in natural odors, $p_i \ll 1/2$, such optimization is equivalent to a partition problem where the N_l probabilities $\{p_i\}$ have to be put into N_r groups (i.e., a group of ligands for each receptor), such that the sum of the elements is close to 1/2, while a minimal number of elements should appear in several groups. Eq. **4** gives the relative cost of violating these two possibly conflicting requirements.

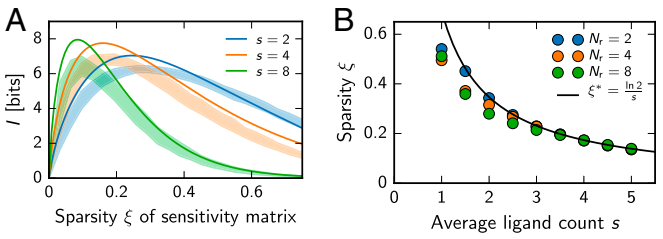

Fig. 2. Receptor arrays with random sensitivity matrices whose sparsity ξ is tuned to match natural statistics achieve near-optimal information transmission of odor composition. (*A*) Information I gained by $N_r = 8$ receptor as a function of the average sparsity ξ of random binary sensitivity matrices for mixtures made of s ligands drawn from a total of $N_l = 32$ ligands. Numerical results (shaded areas; mean ± SD; 32 samples) and analytical results (lines) following from Eq. **4** are shown. (*B*) Sparsity ξ of general binary sensitivity matrices that were numerically optimized for maximal I (symbols) is compared with the prediction from random binary matrices (solid line, Eq. **8**) for different s and N_r at $N_l = 128$.

This partition problem can be solved approximately using random binary sensitivity matrices. The ensemble of such matrices is characterized by a single parameter, the fraction of nonzero entries or sparsity ξ. Fig. 2*A* shows that there is an optimal sparsity ξ^*, at which I is maximized. It follows from $\langle a_n \rangle = 1/2$ that

$$\xi^* \approx \frac{\ln 2}{s}, \qquad \textbf{[8]}$$

where $s = \sum_i p_i$ is the mean mixture size (see *Supporting Information*). This condition for random matrices agrees well with the sparsity found from numerical optimization over all binary matrices (see Fig. 2*B*). However, for small s, the sparsity ξ^* becomes large, which leads to significant correlations $\mathrm{cov}(a_n, a_m)$ and thus reduced performance. Optimal matrices thus have a sparsity that is lower then predicted by Eq. **8** for small mixture sizes s (see Fig. 2*B*).

Wide concentration distributions. In reality, odor concentrations vary widely, and receptor arrays must thus measure both odor composition and concentrations. The concentration of a single ligand can be measured if many receptors react to it with different sensitivities (7). The receptor array is optimal for this task if all possible outputs occur with equal frequency. This is the case if the inverse of the sensitivities follows the same distribution as the ligand concentrations (13), which is known as Laughlin's principle. However, it is not clear how this principle can be generalized for measuring the concentration of multiple ligands simultaneously.

We study this problem by considering random sensitivities that are lognormally distributed. This choice is motivated by the complex interaction between receptors and ligands, which typically leads to normally distributed binding energies (28). We will show later that experimentally measured sensitivities indeed appear to be lognormally distributed. Lognormal distributions are characterized by two parameters, the mean $\overline{S}$ and the SD λ of the underlying normal distribution. We thus next ask how these parameters have to be chosen to maximize the mutual information I. To estimate I, we need to consider the excitations e_n, which approximately also follow a lognormal distribution (29). Their statistics are given by Eq. **6** and read $\langle e_n \rangle_{c,S} = \overline{S} \langle c_{\mathrm{tot}} \rangle$ and $\mathrm{cov}_{c,S}(e_n, e_m) = \overline{S}^2 \mathrm{var}(c_{\mathrm{tot}}) + \delta_{nm} \mathrm{var}(S) \sum_i \langle c_i^2 \rangle$, where $c_{\mathrm{tot}} = \sum_i c_i$ and $\mathrm{var}(S) = \overline{S}^2 [\exp(\lambda^2) - 1]$. We use this to calculate $\langle a_n \rangle$ from Eq. **2** and find that the receptor array is optimal ($\langle a_n \rangle = 1/2$) if (see *Supporting Information*)

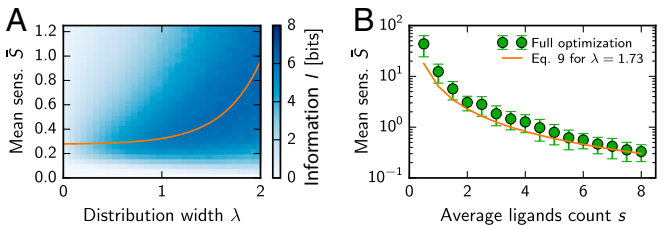

Fig. 3. Random receptor arrays with a suitable mean sensitivity $\bar{S}$ and distribution width λ can transmit information about both odor concentration and composition. (*A*) Information I for lognormally distributed sensitivities as a function of the mean $\bar{S}$ and width λ of the distribution. The shown mean of I was calculated from Eqs. **1**–**3** using Monte Carlo sampling of 32 realizations per point. The orange line marks the optimum given by Eq. **9**. (*B*) Mean sensitivity $\bar{S}$ for different average mixture sizes s. Numerical optimizations over general sensitivity matrices (symbols; mean ± SD; 64 samples) are compared with lognormally distributed matrices (solid line, Eq. **9**) with $\lambda = 1.73$, equal to the mean of the numerical data. Additional parameters in *A* and *B* are the same as in Fig. 1.

$$\bar{S} = \frac{1}{\langle c_{\text{tot}} \rangle} \left[1 + \frac{\text{var}(c_{\text{tot}})}{\langle c_{\text{tot}} \rangle^2} + \frac{\sum_i \langle c_i^2 \rangle}{\langle c_{\text{tot}} \rangle^2} \left(e^{\lambda^2} - 1 \right) \right]^{\frac{1}{2}}. \qquad [9]$$

We test this equation by numerically calculating the mutual information I as a function of $\bar{S}$ and λ. Fig. 3*A* shows that Eq. **9** predicts the optimal parameters of lognormally distributed sensitivities very well. Fig. 3*B* shows that this result also predicts the mean $\bar{S}$ for numerical optimizations over general sensitivity matrices.

Log-normally distributed sensitivities perform badly if the distribution width λ is small (see Fig. 3*A*). This is expected because receptors with narrowly distributed S_{ni} respond similarly to all ligands, leading to large correlations $\text{cov}(a_n, a_m)$ and thus reduced performance I. Interestingly, for large enough λ, the correlations are so small that the exact value of λ does not influence I significantly (see Fig. 3*A*). In fact, for very large λ, the S_{ni} are likely very large or very small compared with $\bar{S}$. When $\bar{S}$ is chosen according to Eq. **9**, receptors can thus only detect whether ligands are present or not, corresponding to the binary sensitivities discussed above, which cannot resolve the concentration of the ligands. Consequently, λ must influence how well such receptor arrays can resolve concentrations.

Trade-off between concentration resolution and mixture discriminability. When the distribution width λ is large, the receptor arrays have similar performance I, so they are equally good at the combined problem of resolving concentrations and discriminating mixtures. However, the performance in the individual problems can vary widely. Because, in many contexts, we might wish to trade off performance, say, by sacrificing some ability to discriminate mixtures in favor of a better concentration resolution, we next investigate these properties in detail.

We define the concentration resolution R as the ratio of the concentration c at which a single ligand is presented and the concentration change δc that is necessary to register a change, $R = c/\delta c$. Here, we consider the simple case where η additional receptors have to be excited to register a change in concentration. R is a function of the concentration c at which it is measured and its maximal value

$$R_{\max} = \frac{N_{\text{r}}}{\sqrt{2\pi}\eta\lambda} \qquad [10]$$

is obtained for $c = \bar{S}^{-1} \exp[(1/2)\lambda^2]$, which is the inverse of the median of the sensitivity distribution (see *Supporting Information*).

The range of concentrations that can be detected by the receptor array is given by the ratio of the largest concentration $c_{\max}$ at which concentration differences can be detected to the lowest detectable concentration $c_{\min}$, the odor detection threshold (30). In terms of η, the logarithm of the concentration range $\zeta = c_{\max}/c_{\min}$ reads (see *Supporting Information*)

$$\ln(\zeta) = \sqrt{8}\lambda \text{erf}^{-1}\left(1 - \frac{2\eta}{N_{\text{r}}}\right), \qquad [11]$$

where $\text{erf}^{-1}(z)$ is the inverse error function. Eq. **11** shows that λ determines the number of concentration decades over which the receptor array is sensitive.

Taken together, λ has opposing effects on the resolution and the range of concentration measurements (see Fig. 4*A*). Consequently, λ can be tuned either for receptors that resolve concentrations well or cover a large concentration range. If only single ligands are measured, the optimal λ only depends on the concentration distribution $P_{\text{env}}(\boldsymbol{c})$. In this case, the mutual information I can be calculated from the resolution function $R(c)$, and optimizing $R(c)$ is equivalent to maximizing I (31). For odor mixtures, I accounts for a combination of the concentration resolution and the mixture discrimination, and maximizing I does not uniquely determine an optimal receptor array. We thus next study how the distribution width λ influences the ability to discriminate mixtures.

We first consider mixtures of s ligands, each at concentration c, and determine the maximal size $s_{\max}$ where adding an additional ligand does not significantly alter the activity pattern. Here $s_{\max}$ is given by the largest s that obeys (see *Supporting Information*)

$$\frac{\text{d}\langle a_n \rangle_S}{\text{d}s} \geq \frac{\eta}{N_{\text{r}}}, \qquad [12]$$

where $\langle a_n \rangle_S \approx 1 - F_{\text{LN}}[c^{-1}; \bar{S}s, \text{var}(S)s]$ with $F_{\text{LN}}(x; \mu, \sigma^2)$ being the cumulative distribution function of a lognormal distribution with mean μ and variance σ^2. Fig. 5*A* shows that $s_{\max}$ increases with decreasing concentrations, but, if the concentration falls below the odor detection threshold, individual ligands cannot be detected (dotted lines).

Not all mixtures with less then $s_{\max}$ ligands can be distinguished from each other. We show this by calculating the Hamming distance h of the activity patterns $\boldsymbol{a}$ of two mixtures, i.e., the number of differences in the output. For simplicity, we consider mixtures that contain s ligands, sharing s_{b} of them. In this case, a given receptor is activated by one of the mixtures if $e_{\text{b}} + e_{\text{d}} > 1$, where e_{b} and e_{d} are the excitations caused by the s_{b} shared and the $s - s_{\text{b}}$ different ligands, respectively. Approximating the probability distribution of the excitations as a lognormal distribution, we can calculate the expected distance h (see *Supporting Information*). Fig. 5*B* shows that this approximation (solid lines) agrees well with numerical calculations (symbols). The figure also shows that mixtures can only be distinguished well if the concentration of the constituents is in the right range. This is because receptors are barely excited for too small concentrations, whereas they are saturated for large concentrations. The distance h also strongly

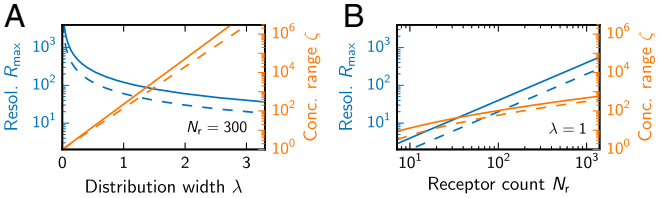

Fig. 4. The width λ of the sensitivity distribution has opposing effects on concentration resolution $R_{\max}$ (blue, Eq. **10**) and range ζ (orange, Eq. **11**). (*A*) $R_{\max}$ and ζ as a function of the width λ for $N_r = 300$ receptors. (*B*) $R_{\max}$ and ζ as a function of N_r for $\lambda = 1$. In *A* and *B*, $\eta = 1$ (solid lines) and $\eta = 2$ (dashed lines) changes in the output pattern are required to distinguish inputs.

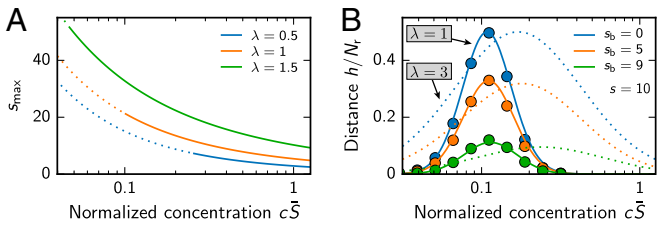

Fig. 5. The discriminability of mixtures strongly depends on the concentrations at which odors are presented. (*A*) Maximal mixture size s_{max} (from Eq. **12**) as a function of the ligand concentration c for different widths λ of the sensitivity distribution at $N_r/\eta = 300$. Dotted lines indicate where c is below the detection threshold for single ligands. (*B*) Mean difference h in the activation pattern of two mixtures of size $s = 10$ as a function of c for different numbers s_b of shared ligands and widths λ. Analytical results (lines) are compared with numerical simulations (symbols).

depends on the number s_b of shared ligands between the two mixtures, which has also been shown experimentally (32). The distance vanishes for $s_b = s$, but Fig. 5*B* shows that a single different ligand can be sufficient to distinguish mixtures in the right concentration range (green line). This range increases with the width λ of the sensitivity distribution, similar to the range over which concentrations can be measured (see Eq. **11**). The suitable concentration range is also a function of the mean sensitivity $\bar{S}$, which, in turn, must be adjusted to the odor statistics (see Eq. **9**). Consequently, our model predicts that only mixtures with total concentrations near the average concentration in natural mixtures can be distinguished well.

Experimentally Measured Receptor Arrays. The response of receptors to individual ligands has been measured experimentally for flies (33) and humans (34). We use these published data to estimate the statistics of realistic sensitivity matrices as described in *Supporting Information*. Fig. 6 shows the histograms of the logarithms of the sensitivities for flies and humans. Both histograms are close to a normal distribution, with similar SDs $\lambda_{exp} \approx 1.1$, which implies lognormally distributed sensitivities. Using a simple binding model between receptors and ligands, λ_{exp} can also be interpreted as the SD of the interaction energies (see *Supporting Information*). Consequently, these interaction energies exhibit a similar variation on the order of 1 k_BT for both organisms, which could be caused by the biophysical similarity of the receptors.

We next use the measured lognormal distribution for the sensitivities to compare the concentration resolution R predicted by Eq. **10** to measured "just noticeable relative differences" R^{-1} (23). For humans ($N_r = 300$), the measured values are as low as 4% (35), which implies $\eta\lambda \approx 4.8$. Using $\lambda \approx 1.1$, this suggest that about four receptors have to be activated until a change in concentration can be registered. Additionally, our theory predicts that humans can sense concentrations over about 2.6 orders of magnitude, which follows from Eq. **11** for $\lambda = 1.1$, $\eta = 1$, and $N_r = 300$. However, we are not aware of any measurements of the concentration range for humans.

Our theory also predicts the maximal number of ligands that can be distinguished as a function of the concentration c of the individual ligands. For $\lambda \approx 1.1$, we expect that the maximal number s_{max} of ligands in a mixture is around 20 if individual ligands can be detected (see Fig. 5*A*). Experimental studies report similar numbers, e.g., $s_{max} \approx 15$ (36) and $s_{max} < 30$ (6). However, Fig. 5*A* shows that s_{max} strongly depends on the concentration of the individual ligands and thus on experimental details. Similarly, how well mixtures can be discriminated also depends strongly on the ligand concentration. Fig. 5*B* shows that the concentration range over which mixtures can be distinguished is less than an order of magnitude for $\lambda \approx 1.1$.

Discussion

We studied how arrays of olfactory receptors can be used to measure odor mixtures, focusing on the combinatorial code of olfaction, i.e., how the combined response of multiple receptors can encode the composition (quality) and the concentration (quantity) of odors. Such arrays are optimal if each receptor responds to half of the encountered odors and the receptors have distinct ligand binding profiles to minimize correlations.

Our simple model of binary receptors can, in principle, distinguish a huge number of odors, because there are $\sim 10^{90}$ different output combinations for $N_r = 300$. However, it is not clear whether all outputs are achievable and how they are used to distinguish odors. We showed that the mean receptor sensitivity must be tailored to the mean concentration to best use the large output space. Another important parameter of receptor arrays is the fraction of receptors that is activated by a single ligand, which is equivalent to the sparsity ξ in the simple case of binary sensitivities. If ξ is small, combining different ligands typically leads to unique output patterns that allow identification of the mixtures, but the concentration of isolated ligands cannot be measured reliably, because only a few receptors are involved. Conversely, if ξ is large, mixtures of multiple ligands will excite almost all receptors, such that neither the odor quality nor the odor quantity can be measured reliably. However, here, the concentration of an isolated ligand can be measured precisely. We discussed this property in detail for sensitivities that are lognormally distributed, where the width λ controls whether mixtures can be distinguished well or concentrations can be measured reliably. Interestingly, experiments find that individual ligands at moderate concentration only excite a few glomeruli (37), but natural odors at native concentrations can excite many (38). This could imply that the sensitivities are indeed adapted such that each receptor is excited about half the time for natural odors.

Our model implies that having more receptor types can improve all properties of the receptor array. In particular, both the concentration resolution R and the typical distance h between mixtures are proportional to N_r, a prediction that can be tested experimentally. For instance, mice, with $N_r \approx 1{,}000$ receptor types, are very good at identifying a single odor in a mixture (39), but flies, with $N_r = 52$ (33), should perform much worse. However, quantitative comparisons might be difficult because the discrimination performance strongly depends on the normalized concentration $c\bar{S}$ at which odors are presented. In fact, we predict that mixtures can hardly be distinguished if the concentration of the individual ligands is changed by an order of magnitude (see Fig. 5*B*).

Our results also apply to artificial chemical sensor arrays known as "artificial noses" (40, 41). Having more sensors improves the general performance of the array, but it is also important to tune the sensitivity of individual sensors. Here, sensors should be as diverse as possible while still responding to about half the incoming mixtures. Unfortunately, building such chemical sensors is difficult, and their binding properties are hard to control (41). If

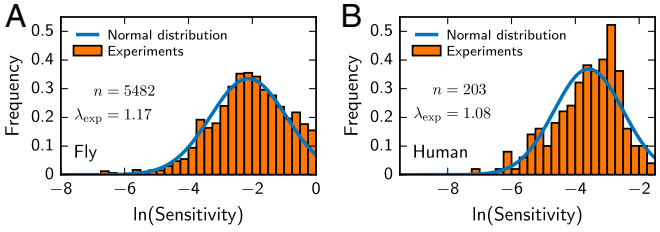

Fig. 6. Sensitivities of olfactory receptors appear to be lognormally distributed for (*A*) flies (33) and (*B*) humans (34). The histograms of the logarithms of n entries of the sensitivity matrix (orange) are compared with a normal distribution (blue) with the same mean and SD λ_{exp}.

the sensitivity matrix of the sensor array is known, our theory can be used to estimate the information I_n that receptor n contributes as $I_n \approx H(\langle a_n \rangle) + H(1 - \langle a_n \rangle) - (4/\ln 2)\sum_{m \neq n} \mathrm{cov}(a_n, a_m)^2$ where $H(p) = -p \log_2 p$, such that $I = \sum_n I_n$ (see Eq. **4**). This can then be used for identifying poor receptors that contribute only a little information to the overall results.

Our focus on the combinatorial code of the olfactory system certainly neglects intricate details of the system. For instance, we do not consider the dynamics of sniffing and odor absorption, which are the first processing steps and influence the perception (42). Further, our simple model of the binding of odorants to receptors, described by sensitivity matrices with independent entries, neglects biophysical constraints that will cause chemically similar ligands to excite similar receptors (8, 43). This is important because it makes it difficult to distinguish similar ligands (44), and it might thus be worthwhile to dedicate more receptors to such a part of chemical space. Additionally, receptors or glomeruli might interact with each other, e.g., causing inhibition reducing the signal upon binding a ligand (45). We can, in principle, discuss inhibition in our model by allowing for negative sensitivities, but more complicated features cannot be captured by the linear relationship in Eq. **1**. One important nonlinearity is the dose–response curve of individual receptor neurons (21), which we approximate by a step function (see Eq. **2**). This simplification reduces the information capacity of a single glomerulus to 1 bit, whereas it is likely higher in reality. However, we expect that allowing for multiple output levels would only increase the concentration resolution and not change the discriminability of mixtures very much (23). Additionally, these perceptual quantities could be influenced by other processes, e.g., lateral inhibition between glomeruli (11, 46) and top-down modulation that adjusts the sensory system based on behavior (46). Besides such enhancements of olfactory sensing, further processing can only remove information, so our results provide an upper bound for the ability to recognize odors.

ACKNOWLEDGMENTS. We thank Carl Goodrich, Venkatesh N. Murthy, and Michael Tikhonov for helpful discussions and a critical reading of the manuscript. This research was funded by the National Science Foundation (NSF) through DMR-1435964, DMR-1420570, and DMS-1411694. M.P.B. is an investigator of the Simons Foundation. D.Z. was also funded by the German Science Foundation through ZW 222/1-1, the NSF through PHY11-25915, the National Institutes of Health Award 5R25GM067110-07, and the Moore Foundation Award 2919.

1. Touhara K, Vosshall LB (2009) Sensing odorants and pheromones with chemosensory receptors. *Annu Rev Physiol* 71:307–332.
2. Su CY, Menuz K, Carlson JR (2009) Olfactory perception: Receptors, cells, and circuits. *Cell* 139(1):45–59.
3. Mainland JD, Lundström JN, Reisert J, Lowe G (2014) From molecule to mind: An integrative perspective on odor intensity. *Trends Neurosci* 37(8):443–454.
4. Verbeurgt C, et al. (2014) Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PLoS One* 9(5):e96333.
5. Dunkel M, et al. (2009) SuperScent—A database of flavors and scents. *Nucleic Acids Res* 37(Database issue):D291–D294.
6. Weiss T, et al. (2012) Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. *Proc Natl Acad Sci USA* 109(49):19959–19964.
7. Hopfield JJ (1999) Odor space and olfactory processing: Collective algorithms and neural implementation. *Proc Natl Acad Sci USA* 96(22):12506–12511.
8. Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial receptor codes for odors. *Cell* 96(5):713–723.
9. Hasin Y, et al. (2008) High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet* 4(11):e1000249.
10. Maresh A, Rodriguez Gil D, Whitman MC, Greer CA (2008) Principles of glomerular organization in the human olfactory bulb—Implications for odor processing. *PLoS One* 3(7):e2640.
11. Murthy VN (2011) Olfactory maps in the brain. *Annu Rev Neurosci* 34:233–258.
12. Leon M, Johnson BA (2003) Olfactory coding in the mammalian olfactory bulb. *Brain Res Brain Res Rev* 42(1):23–32.
13. Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch C* 36(9-10):910–912.
14. Ruderman DL, Bialek W (1994) Statistics of natural images: Scaling in the woods. *Phys Rev Lett* 73(6):814–817.
15. Lewicki MS (2002) Efficient coding of natural sounds. *Nat Neurosci* 5(4):356–363.
16. Tkacik G, Callan CG, Jr, Bialek W (2008) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci USA* 105(34):12265–12270.
17. Hummel R (1977) Image enhancement by histogram transformation. *Comput Graph Image Process* 6(2):184–195.
18. Wright GA, Thomson MG (2005) Odor perception and the variability in natural odor scenes. *Integrative Plant Biochemistry*, Recent Advances in Phytochemistry (Elsevier, New York), Vol 39, pp 191–226.
19. McGann JP, et al. (2005) Odorant representations are modulated by intra- but not interglomerular presynaptic inhibition of olfactory sensory neurons. *Neuron* 48(6): 1039–1053.
20. Lin Y, Shea SD, Katz LC (2006) Representation of natural stimuli in the rodent main olfactory bulb. *Neuron* 50(6):937–949.
21. Reisert J, Restrepo D (2009) Molecular tuning of odorant receptors and its implication for odor signal processing. *Chem Senses* 34(7):535–545.
22. Lowe G, Gold GH (1995) Olfactory transduction is intrinsically noisy. *Proc Natl Acad Sci USA* 92(17):7864–7868.
23. Koulakov A, Gelperin A, Rinberg D (2007) Olfactory coding with all-or-nothing glomeruli. *J Neurophysiol* 98(6):3134–3142.
24. Stevens CF (2015) What the fly's nose tells the fly's brain. *Proc Natl Acad Sci USA* 112(30):9460–9465.
25. Knudsen JT, Tollsten L, Bergström LG (1993) Floral scents? A checklist of volatile compounds isolated by head-space techniques. *Phytochemistry* 33(2):253–280.
26. Atick JJ (2011) Could information theory provide an ecological theory of sensory processing? *Network* 22(1-4):4–44.
27. Sessak V, Monasson R (2009) Small-correlation expansions for the inverse ising problem. *J Phys A* 42:055001.
28. Lancet D, Sadovsky E, Seidemann E (1993) Probability model for molecular recognition in biological receptor repertoires: Significance to the olfactory system. *Proc Natl Acad Sci USA* 90(8):3715–3719.
29. Fenton LF (1960) The sum of log-normal probability distributions in scatter transmission systems. *IRE Trans Commun Syst* 8(1):57–67.
30. Abraham MH, Sánchez-Moreno R, Cometto-Muñiz JE, Cain WS (2012) An algorithm for 353 odor detection thresholds in humans. *Chem Senses* 37(3):207–218.
31. Bialek W (2012) *Biophysics: Searching for Principles* (Princeton Univ Press, Princeton, NJ).
32. Bushdid C, Magnasco MO, Vosshall LB, Keller A (2014) Humans can discriminate more than 1 trillion olfactory stimuli. *Science* 343(6177):1370–1372.
33. Münch D, Galizia CG (2016) Door 2.0—Comprehensive mapping of *Drosophila melanogaster* odorant responses. *Sci Rep* 6:21841.
34. Mainland JD, Li YR, Zhou T, Liu WLL, Matsunami H (2015) Human olfactory receptor responses to odorants. *Sci Data* 2:150002.
35. Cain WS (1977) Differential sensitivity for smell: "Noise" at the nose. *Science* 195(4280):796–798.
36. Jinks A, Laing DG (1999) A limit in the processing of components in odour mixtures. *Perception* 28(3):395–404.
37. Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD (2009) Odor coding by a Mammalian receptor repertoire. *Sci Signal* 2(60):ra9.
38. Vincis R, Gschwend O, Bhaukaurally K, Beroud J, Carleton A (2012) Dense representation of natural odorants in the mouse olfactory bulb. *Nat Neurosci* 15(4):537–539.
39. Rokni D, Hemmelder V, Kapoor V, Murthy VN (2014) An olfactory cocktail party: Figure-ground segregation of odorants in rodents. *Nat Neurosci* 17(9):1225–1232.
40. Albert KJ, et al. (2000) Cross-reactive chemical sensor arrays. *Chem Rev* 100(7): 2595–2626.
41. Stitzel SE, Aernecke MJ, Walt DR (2011) Artificial noses. *Annu Rev Biomed Eng* 13:1–25.
42. Wachowiak M (2011) All in a sniff: Olfaction as a model for active sensing. *Neuron* 71(6):962–973.
43. Hallem EA, Carlson JR (2006) Coding of odors by a receptor repertoire. *Cell* 125(1): 143–160.
44. Perez M, Giurfa M, d'Ettorre P (2015) The scent of mixtures: Rules of odour processing in ants. *Sci Rep* 5:8659.
45. Ukhanov K, Corey EA, Brunert D, Klasen K, Ache BW (2010) Inhibitory odorant signaling in Mammalian olfactory receptor neurons. *J Neurophysiol* 103(2):1114–1122.
46. Wilson RI (2013) Early olfactory processing in *Drosophila*: Mechanisms and principles. *Annu Rev Neurosci* 36:217–241.
47. Hansen N (2006) The CMA evolution strategy: A comparing review. *Towards a New Evolutionary Computation*, eds Lozano JA, Larrañaga P, Inza I, Bengoetxea E (Springer, New York), pp 75–102.