

# Olfactory Navigation and the Receptor Nonlinearity

 Jonathan D. Victor,<sup>1</sup>  Sebastian D. Boie,<sup>1</sup>  Erin G. Connor,<sup>2</sup>  John P. Crimaldi,<sup>2</sup> G. Bard Ermentrout,<sup>3</sup> and  Katherine I. Nagel<sup>4</sup>

<sup>1</sup>Feil Family Brain and Mind Research Institute, Weill Cornell Medical College, New York, New York 10065, <sup>2</sup>Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, Colorado 80309, <sup>3</sup>Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, and <sup>4</sup>Neuroscience Institute, New York University Langone Medical Center, New York, New York 10016

The demands on a sensory system depend not only on the statistics of its inputs but also on the task. In olfactory navigation, for example, the task is to find the plume source; allocation of sensory resources may therefore be driven by aspects of the plume that are informative about source location, rather than concentration per se. Here we explore the implications of this idea for encoding odor concentration. To formalize the notion that sensory resources are limited, we considered coding strategies that partitioned the odor concentration range into a set of discriminable intervals. We developed a dynamic programming algorithm that, given the distribution of odor concentrations at several locations, determines the partitioning that conveys the most information about location. We applied this analysis to planar laser-induced fluorescence measurements of spatiotemporal odor fields with realistic advection speeds (5–20 cm/s), with or without a nearby boundary or obstacle. Across all environments, the optimal coding strategy allocated more resources (i.e., more and finer discriminable intervals) to the upper end of the concentration range than would be expected from histogram equalization, the optimal strategy if the goal were to reconstruct the plume, rather than to navigate. Finally, we show that ligand binding, as captured by the Hill equation, transforms odorant concentration into response levels in a way that approximates information maximization for navigation. This behavior occurs when the Hill dissociation constant is near the mean odor concentration, an adaptive set-point that has been observed in the olfactory system of flies.

**Key words:** adaptation; dynamic programming; efficient coding; Hill equation; information theory

## Significance Statement

The first step of olfactory processing is receptor binding, and the resulting relationship between odorant concentration and the bound receptor fraction is a saturating one. While this Hill nonlinearity can be viewed as a distortion that is imposed by the biophysics of receptor binding, here we show that it also plays an important information-processing role in olfactory navigation. Specifically, by combining a novel dynamic-programming algorithm with physical measurements of turbulent plumes, we determine the optimal strategy for encoding odor concentration when the goal is to determine location. This strategy is distinct from histogram equalization, the strategy that maximizes information about plume concentration, and is closely approximated by the Hill nonlinearity when the binding constant is near the ambient mean.

## Introduction

The notion of “efficient coding,” that sensory systems exploit the statistical regularities of their natural inputs (Barlow, 1961), has

provided a foundation for understanding their design principles at algorithmic and mechanistic levels. While at first glance it might appear that evolutionary pressure makes efficient coding unavoidable, two aspects of this principle are decidedly nontrivial: (1) that natural inputs indeed possess substantial statistical regularity; and (2) that this regularity can be exploited without substantially compromising their use for decision or action. This second aspect is often overlooked, as it is implicitly assumed that creating an internal representation of the sensory environment is an adequate surrogate for their ultimate use. Thus, information-theoretic measures of the fidelity of this representation are taken as the starting point for normative theories of sensory processing

Received Sept. 27, 2018; revised Jan. 29, 2019; accepted Feb. 23, 2019.

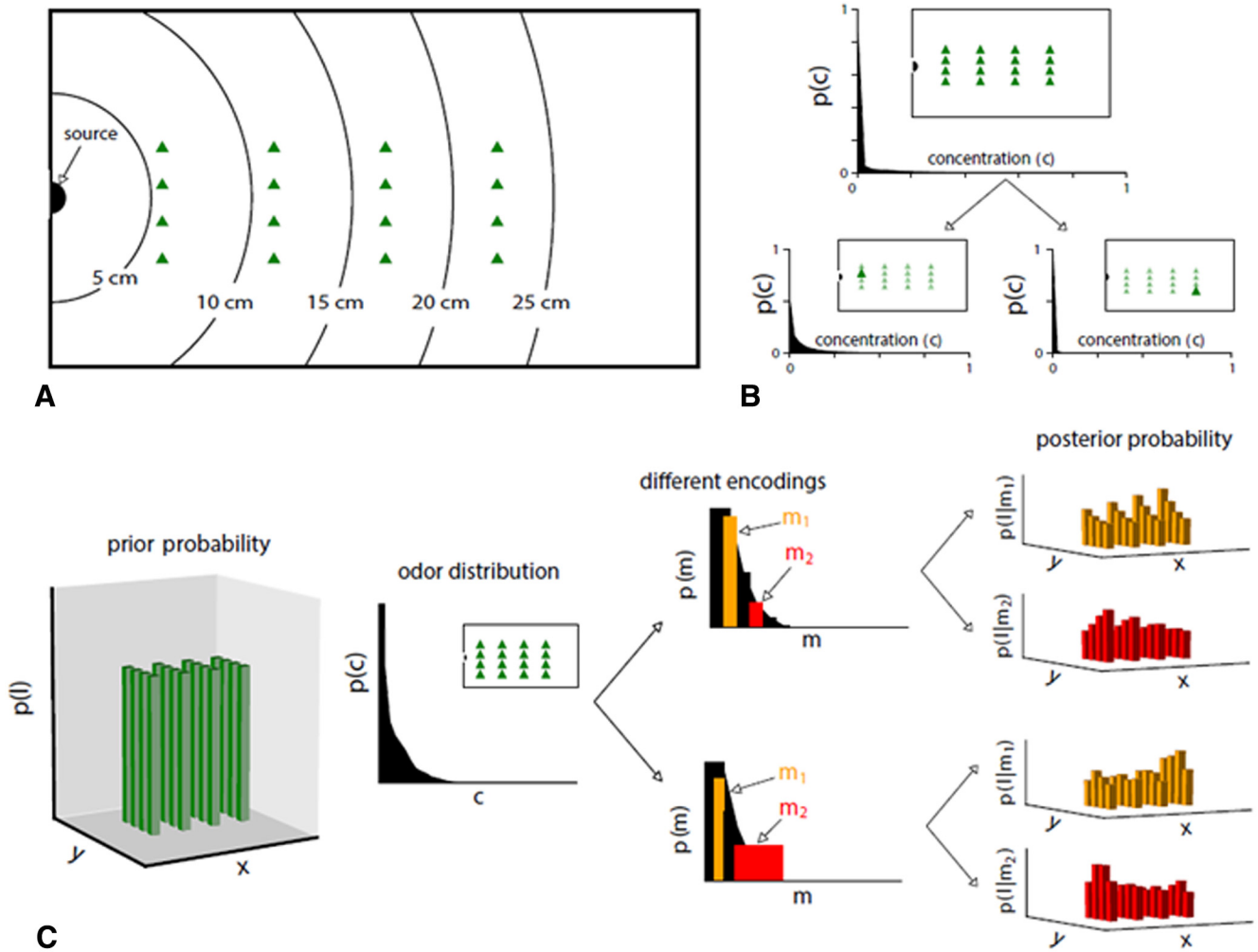
Author contributions: J.D.V., J.P.C., and K.I.N. designed research; J.D.V., S.D.B., E.G.C., and J.P.C. performed research; J.D.V., S.D.B., E.G.C., and J.P.C. analyzed data; J.D.V. wrote the first draft of the paper; J.D.V., S.D.B., E.G.C., J.P.C., G.B.E., and K.I.N. edited the paper.

S.D.B. and J.D.V. were supported by National Science Foundation Grant IOS 1555891. E.G.C. and J.P.C. were supported by National Science Foundation Grant PHY 1555862. G.B.E. was supported by National Science Foundation Grant PHY 1555916. K.I.N. was supported by National Science Foundation Grant IOS 1555933. These grants were all part of a collaborative research project supported by the NSF IdeasLab initiative. We thank Alex Dimitrov and Ramin Zabih for helpful comments.

The authors declare no competing financial interests.

Correspondence should be addressed to Jonathan D. Victor at [jdvicto@med.cornell.edu](mailto:jdvicto@med.cornell.edu).

<https://doi.org/10.1523/JNEUROSCI.2512-18.2019>  
Copyright © 2019 the authors



**Figure 1.** An overview of the approach. **A**, Odor concentration is measured at multiple grid locations (triangles) within a dynamic odor plume. **B**, Since the odor concentration varies with time, each location yields a distribution of odor concentrations. Large histogram represents the distribution of odor concentrations across all grid points. The two smaller histograms represent the distribution of odor concentrations at two example grid points. At the grid point close to the odor source (bottom left histogram), high odor concentrations are more frequent than at a grid point far from the source (bottom right histogram). **C**, Analysis of alternative schemes for encoding odor concentration. As diagrammed on the left, locations across the grid points are assigned equal *a priori* probability. An odor sample is obtained at a randomly chosen grid point, and encoded into a code word that represents a specific bin of odor concentrations (middle). Then, for each code word, the *a posteriori* probability of location within the grid is computed via Bayes Theorem (right). The dependence of the *a posteriori* probability distribution on the code word drives a computation of the Shannon mutual information between location and code word. We determine the encoding scheme (i.e., the partitioning used for encoding) that maximizes this information, given a fixed number of code words. Modified with permission from Boie et al. (2018).

and a way to understand the role(s) of neural mechanisms (Barlow, 1961; Laughlin, 1981; Srinivasan et al., 1982; Olshausen and Field, 1996; Karklin and Lewicki, 2009; Tkacik et al., 2010; Hermundstad et al., 2014).

Olfactory navigation, however, is a critical and pervasive animal behavior in which the end-use of sensory information and the construction of an internal representation may be distinct. Specifically, an animal attempting to locate the source of an odor may create a map as an intermediate step (Vergassola et al., 2007; Gagliardo, 2013; Grünbaum and Willis, 2015; Jacobs et al., 2015), but successful navigation can also be based on an entirely local decision strategy (Pierce-Shimomura et al., 1999; Wen et al., 2012; Schulze et al., 2015). From a normative viewpoint, a system whose goal is to construct an internal representation of the olfactory environment should be optimized to transmit information about odor concentration, but a system whose goal is olfactory navigation should be optimized to transmit information relevant to locating the odor source. For example, because odor plumes are turbulent (Celani et al., 2014; Connor et al., 2018), “hits” of

high concentrations, though rare, may be valuable localizing cues, so a sensory system optimized for navigation may devote disproportionate resources to encoding high concentrations.

With this in mind, we seek to determine the optimal encoding strategy for a sensory system whose goal is olfactory navigation. To keep the focus on the strategy, rather than the concrete form of the code, we use information theory as a framework. We postulate (Fig. 1) that the entire odor concentration range is encoded into “code words,” each representing a nonoverlapping interval. (We use the term “code word” to emphasize that our focus is on allocation strategy, not the specific form of the code. For concreteness, each code word may be thought of as a particular number of spikes, but this is not a necessity for the analysis.) We then determine the choice of intervals that maximizes the information conveyed about source location. This optimal partitioning indicates how coding resources are allocated: greater resources to odor ranges that are subdivided into more code words.

If the goal were to maximize the information conveyed about odor concentration, the optimal strategy is histogram equaliza-

tion (HE) (Laughlin, 1981; Tadmor and Tolhurst, 2000): intervals are chosen so that each code word is used equally often. Because odor concentration distributions have a strong positive skew, HE allocates most code words to the low end of the concentration range. This corresponds to applying a highly compressive nonlinearity to odor concentration, and then linear encoding of this nonlinearity's output.

As we will show, the optimal strategy for transmitting information about location is not HE, but rather, one that devotes a greater number of code words to the high end of the concentration range. We identify these optimal codes via a novel dynamic programming algorithm, driven by physical measurements (planar laser-induced fluorescence) (Connor et al., 2018) of the spatiotemporal concentration distribution in real plumes.

While the optimization algorithm is highly nonbiological, we show that the efficiency that it achieves is nearly matched by a biophysical mechanism that is already known to be present (Nagel and Wilson, 2011; Gorur-Shandilya et al., 2017). Receptor binding, the first stage of olfaction, implies a gently saturating relationship between odorant concentration and the fraction of bound receptors. With the binding constant set to the ambient mean (as suggested by studies of adaptation) (Cao et al., 2016; Gorur-Shandilya et al., 2017), this Hill nonlinearity, followed by linear encoding of the bound fraction, leads to an allocation of code words that closely approximates optimal performance. Thus, receptor binding is not only a necessary step in olfactory processing, but one whose nonlinear characteristics match the needs of navigation.

## Materials and Methods

This work combines experimental measurements of spatiotemporal plumes with a novel algorithm for determining optimal coding strategies.

**Experimental methods.** Spatiotemporal distribution of odorant concentrations was determined by planar laser-induced fluorescence at a frame rate of 15 Hz and a spatial resolution of 0.74 mm in five environments; these measurements served as the starting point for information-theoretic calculations. Measurement methods, detailed by Connor et al. (2018), are identical to those of Boie et al. (2018) and are provided here for the reader's convenience.

**Plume measurements.** Apparatus consisted of a low-speed air wind tunnel with a test section of cross-section  $0.3 \times 0.3$  m, and 1 m in length, into which an odor surrogate (acetone) was released. Turbulence is induced in the tunnel by an entrance grid consisting of 6.4-mm-diameter rods and a 25.5 mm mesh spacing, followed by a 15-cm-long honeycomb section. Acetone is isokinetically released into flowing air through a 9.5-mm-diameter tube on the tunnel centerline, whose orifice was 10 cm downstream of the turbulence grid.

Acetone vapor, used as an odor surrogate, was generated by bubbling a carrier gas, consisting of air (59% v/v) and helium (41% v/v) through liquid acetone, to produce neutral buoyancy in the wind tunnel. Wind tunnel temperature was stabilized at ambient conditions with a water bath.

Fluorescence was induced with a 1-mm-thick light sheet from a Nd:YAG 266 nm pulsed laser, which illuminated the odor plume in the test section and entered the tunnel through longitudinal slits in its sides. Plume fluorescence, proportional to acetone concentration, was imaged through a glass window in the tunnel using a high quantum efficiency sCMOS camera. The camera had a bit depth of 16, and its framerate, 15 Hz, was synchronized to the laser pulses. Raw images were binned to  $512 \times 512$  pixels, corresponding to a spatial resolution of 0.74 mm/pixel. They were corrected for background and light field inhomogeneities by the following:

$$c(x, y, t) = \frac{1}{a_c} \frac{I(x, y, t)}{F(x, y)} \quad (1)$$

where  $c(x, y, t)$  is the normalized concentration value,  $I(x, y, t)$  is the raw pixel value with background subtracted,  $F(x, y)$  is the flatfield image also with background subtracted, and  $a_c$  is a scaling coefficient that sets  $c = 1$  to correspond to the concentration of acetone at its entrance into the test section.

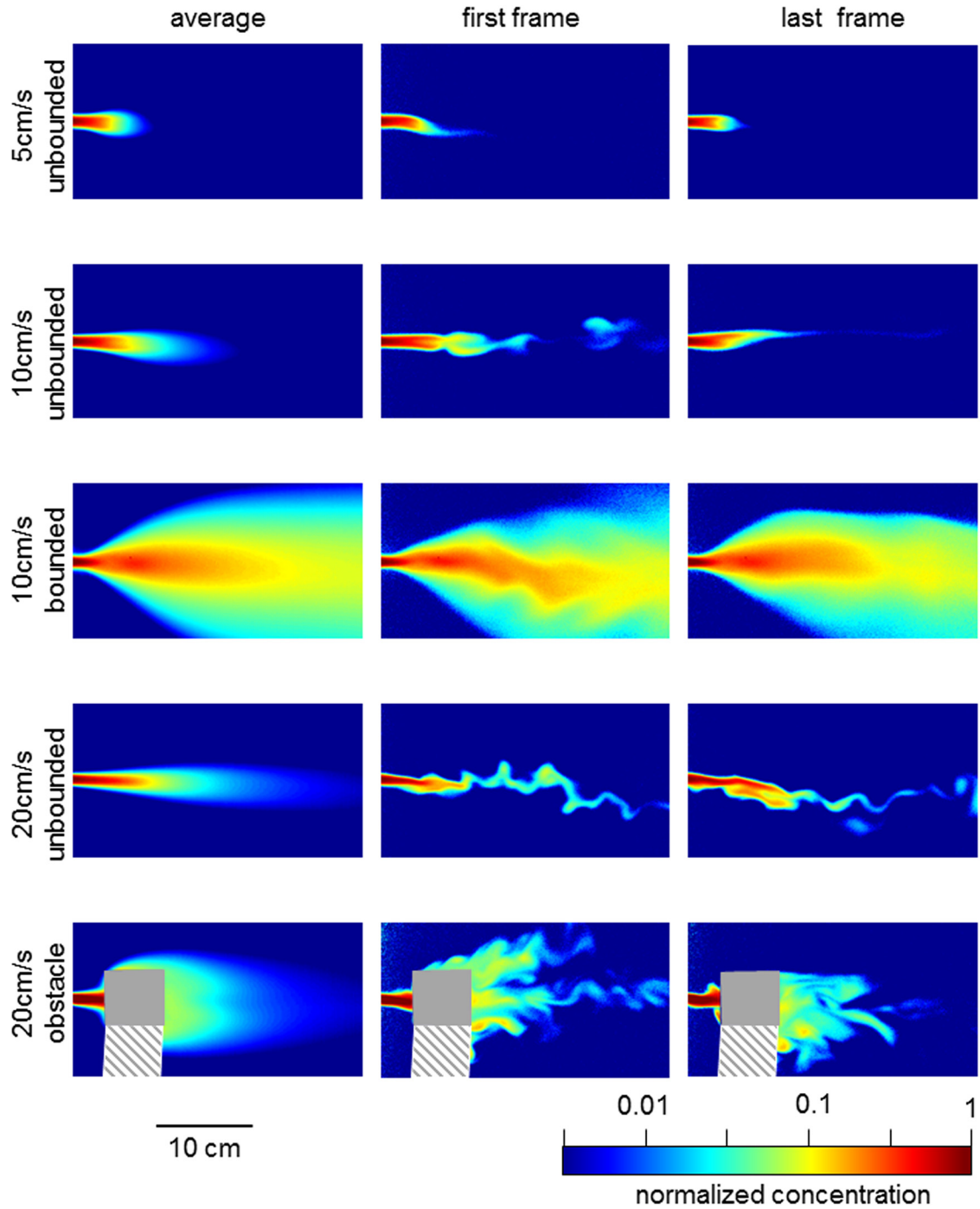
**Olfactory environments.** Planar laser-induced fluorescence measurements were obtained under five conditions ("olfactory environments"): the three described by Boie et al. (2018), referred to here as 5 cm/s unbounded, 10 cm/s unbounded, and 10 cm/s bounded, and two additional ones, referred to here as 20 cm/s unbounded and 20 cm/s obstacle (Fig. 2). In all datasets, acetone was released at the plume centerline. For the "unbounded" datasets, the full cross-section of the wind tunnel was used. For the "bounded" dataset, a false floor was placed immediately below the release point, and extended the length and width of the test section. For the "obstacle" dataset, an obstacle was positioned as shown in Figure 2 bottom row (and there was no false floor). In the 5 cm/s and 10 cm/s environments, 36–40 min of data (32,400–36,000 frames at 15 Hz) were obtained; in the 20 cm/s environments, 20 min of data (18,000 frames at 15 Hz) were obtained. In all environments, the field of view of the camera was 30 cm in the direction of flow and 16 cm wide, and data collection was subdivided into runs of 4 min (3600 frames) each.

**Sampling grids.** In each environment, we computed mutual information between the odorant distribution and location for several grids that sampled the full spatiotemporal distributions obtained above. We focus on three two-dimensional grids: two of these ("wide" and "narrow") had  $4 \times 4$  sample points, identical to those used in Boie et al. (2018); the third ("full") covered the regions of both of these grids and had  $7 \times 7$  sample points. We also analyzed six one-dimensional grids: three grids that were orthogonal to the bulk flow direction and three that were parallel to the bulk flow direction. These spanned the region covered by the "full" two-dimensional grid, and each had 16 sample locations. For the 20 cm/s obstacle environment, grid points that were within the obstacle or in the region of the plume that could not be imaged because of the obstacle were excluded from the analysis.

These grids do not imply an assumption about the navigation strategy; they are merely ways of quantifying the extent to which odor distributions at different locations can be distinguished. The analysis does not depend on the order in which the points are sampled or their coordinates (or whether the coordinates are expressed in a Cartesian or polar system), and does not imply that the navigator uses a one-dimensional or a two-dimensional map. We merely use a variety of grids to ensure that our conclusions are robust. The optimization algorithm itself is independent of the geometry of the grid, and also makes no assumption about the sequence in which the points are sampled.

**Computational methods.** Our computational task is to determine how to encode the range of odor concentrations into a set of code words, so that these code words maximize the mutual information between a single odor sample and the location within a plume at which the sample is obtained. This problem, a combinatorial optimization, has an efficient solution described below. The reason that this efficiency is possible is that the optimization can be cast into a form in which the optimal encoding of the entire odor range can be built up from optimal encodings of smaller ranges. This recasting is a consequence of the "chain rule for entropy" (Cover and Thomas, 1991). As we show, recasting the problem enables a dynamic programming algorithm, very much along the lines of the classic "segmented least-squares" algorithm for finding the best piecewise linear approximation to a function (Kleinberg and Tardos, 2006). We first detail the construction of the algorithm, and then its application to plume measurements. All computations were performed in MATLAB (The MathWorks).

**The algorithm.** To begin, we specify (as in Boie et al., 2018) that the range of odor concentrations is to be encoded into a set of code words  $\mathbf{M} = \{1, \dots, M\}$ , where each code word  $m$  corresponds to an interval of odor concentrations (i.e., a "bin"). We may assume that odor concentration is represented by a standardized variable  $X$  whose range is 0 to 1. A set of code words corresponds to a partitioning of this  $[0, 1]$  range according to a set of bin boundaries  $\mathbf{B} = \{b_0, \dots, b_M\}$ , where  $b_0 = 0$ ,  $b_m \leq b_{m+1}$ , and  $b_M = 1$ . That is, the  $m$ th code word corresponds to an odor concentration from  $b_{m-1}$  to  $b_m$ .



**Figure 2.** The five olfactory environments. Heatmaps represent average odor concentration (first column), and snapshots of odor concentration on the first and last frames of data collection (last two columns). The color scale is logarithmic and covers a concentration ranging from 1 (equal to the inlet concentration) down to 0.003. For the 20 cm/s obstacle dataset: Gray square in the plume centerline represents the obstacle. Hatched parallelogram below the centerline represents the region of the plume that could not be imaged.

As is standard, the mutual information between the code words  $M = \{1, \dots, M\}$  and a set of sample locations  $L = \{1, \dots, L\}$  is given by the following:

$$I(L, M) = H(L) - \sum_{m \in M} p(m)H(L|\{m\}) \quad (2)$$

Here,  $H(L)$  is the entropy of the *a priori* distribution of locations:

$$H(L) = - \sum_{l \in L} p(l) \log_2 p(l) \quad (3)$$

where  $p(l)$  is the *a priori* probability that the navigator is at a location  $l \in L$ . As in Boie et al. (2018), we assume that each of these locations is equally likely *a priori*, so  $p(l) = 1/L$ .  $H(L|\{m\})$  is the entropy of the *a posteriori* distribution of locations, given observation of the code word  $m$ :

$$H(L|\{m\}) = - \sum_{l \in L} p(l|m) \log_2 p(l|m) \quad (4)$$

where  $p(l|m)$  is the conditional probability that the navigator is at a location  $l$  given an observation of an odor sample corresponding to the

code word  $m$  (i.e., in the range  $b_{m-1}$  to  $b_m$ ). This conditional probability can be obtained from the measured odor distribution at each location,  $p(m|l)$ , by Bayes' rule.

Our task is to find the set of boundaries  $\mathbf{B}$  that maximizes Equation 2, given a specific number  $M$  of code words. First, since mutual information is symmetric in its arguments, Equation 2 can also be written as follows:

$$I(\mathbf{M}, \mathbf{L}) = H(\mathbf{M}) - \sum_{l \in \mathbf{L}} p(l) H(\mathbf{M} | \{l\}) \quad (5)$$

where  $H(\mathbf{M})$  is the entropy of the code word distribution across all locations, and  $H(\mathbf{M} | \{l\})$  is the entropy of the code word distribution at the single location  $l \in \mathbf{L}$ .

To maximize Equation 5, we consider a somewhat more general quantity, the entropy of a subset of code words  $\mathbf{M}' \subseteq \mathbf{M}$  observed at a subset of locations  $\mathbf{L}' \subseteq \mathbf{L}$ . This is defined by the following:

$$H(\mathbf{M}' | \mathbf{L}') = - \sum_{m \in \mathbf{M}'} \frac{p(m | \mathbf{L}')}{p(\mathbf{M}' | \mathbf{L}')} \log_2 \frac{p(m | \mathbf{L}')}{p(\mathbf{M}' | \mathbf{L}')} \quad (6)$$

As Equation 6 indicates, if the subset  $\mathbf{M}'$  does not include all of  $\mathbf{M}$ , then samples that are not encoded into one of the code words of  $\mathbf{M}'$  are ignored. While we will make use of Equation 6 for many choices of  $\mathbf{M}'$ , we only need it for two choices of  $\mathbf{L}'$ : when  $\mathbf{L}'$  is just one location ( $\mathbf{L}' = \{l\}$ ), and when it is the complete set of locations  $\mathbf{L}$ , in which case  $H(\mathbf{M}' | \mathbf{L}) = H(\mathbf{M}')$ . Thus, the quantity to be optimized, Equation 5, is a sum of terms, each of which is in the form of Equation 6.

The chain rule for entropy (Cover and Thomas, 1991) provides a way to compute  $H(\mathbf{M}' | \mathbf{L}')$  from a decomposition of  $\mathbf{M}'$  into subsets. It states that the entropy of distribution can be computed in two stages: first, by decomposing it into two components and computing the entropy of that decomposition, and then adding the entropies of the components. Applied to Equation 6, this yields the following:

$$H(\mathbf{M}_u \cup \mathbf{M}_v | \mathbf{L}') = -(p_u \log p_u + p_v \log p_v) + p_u H(\mathbf{M}_u | \mathbf{L}') + p_v H(\mathbf{M}_v | \mathbf{L}') \quad (7)$$

where  $\mathbf{M}_u$  and  $\mathbf{M}_v$  are disjoint subsets, and  $p_u$  and  $p_v$  are the conditional probabilities that a code word will be in  $\mathbf{M}_u$  or  $\mathbf{M}_v$ , given that it is in their union:

$$p_u = \frac{p(\mathbf{M}_u | \mathbf{L}')}{p(\mathbf{M}_u | \mathbf{L}') + p(\mathbf{M}_v | \mathbf{L}')} \text{ and } p_v = \frac{p(\mathbf{M}_v | \mathbf{L}')}{p(\mathbf{M}_u | \mathbf{L}') + p(\mathbf{M}_v | \mathbf{L}')}$$

Since Equation 7 applies to components that sum up to the quantity that we optimize (Eq. 5), it implies that any optimal allocation of code words is built out of smaller optimal solutions. More precisely, if a set of code words  $\mathbf{M}'$  encodes a range of odor concentration values in a way that maximizes the mutual information about location (among all segmentations into  $\mathbf{M}'$  code words), and if  $\mathbf{M}' = \mathbf{M}_u \cup \mathbf{M}_v$  is a partition of  $\mathbf{M}'$  into disjoint subsets, then both  $\mathbf{M}_u$  and  $\mathbf{M}_v$  encode their ranges in a way that maximizes the mutual information about location (among all segmentations into  $\mathbf{M}_u$  and  $\mathbf{M}_v$  code words, respectively). This is exactly the decomposition that enables the “segmented least-squares” algorithm: the best piecewise approximation to a function over a large range is always composed of piecewise approximations that are optimal over smaller ranges (Kleinberg and Tardos, 2006).

We exploit this decomposition as follows. The optimal encoding for the entire range  $[0, 1]$  into  $M$  code words can be broken down into an optimal encoding of the range  $[0, b_{M-1}]$  into  $M-1$  code words, along with the identification of the cutpoint  $b_{M-1}$ , and the assignment of the  $M$ th code word to the range  $[b_{M-1}, 1]$ . Similarly, the optimal encoding of the range  $[0, b_{M-1}]$  into  $M-1$  code words can be broken down into an encoding of the range  $[0, b_{M-2}]$  into  $M-2$  code words, along with the identification of the cutpoint  $b_{M-2}$ , and the assignment of the  $(M-1)$ th code word to the range  $[b_{M-2}, b_{M-1}]$ .

As is typical for dynamic programming algorithms, and in close analogy with the “segmented least-squares” algorithm, this recursion is conveniently implemented in a forward fashion, shown in Figure 3. The

initial step (Fig. 3, left) is to create a library of segmentations of the concentration range into two code words, and to compute the amount of mutual information for each. The library of segments covers not only the complete  $[0, 1]$  range, but also the subranges  $[0, x]$ , where  $x$  samples the range with resolution  $1/N$ . (Here, we use  $N = 256$ , and take the values  $x$  to cover equal quantiles; Fig. 3 diagrams the analysis for  $N = 8$ .) The library contains  $N(N-1)/2$  elements: one for each way of choosing the endpoint  $x$  and a cutpoint  $x'$  in  $[0, x]$ . As this library is exhaustive, it is guaranteed to contain the optimal encoding of each subrange  $[0, x]$  into two code words.

Each subsequent step takes a library that is guaranteed to contain the optimal encoding of each  $[0, x]$  interval into  $R$  code words, and creates a library that is guaranteed to contain the optimal encoding of this interval into  $R+1$  code words. This is done by subdividing each  $[0, x]$  into  $[0, x']$  and  $[x', x]$ , computing mutual information via Equation 7, and identifying the cutpoint  $x'$  that produces the maximum value. In applying Equation 7,  $\mathbf{M}_u$  is the set of  $R$  code words assigned to the interval  $[0, x']$ , and  $\mathbf{M}_v$  is a set containing just one code word, for  $[x', x]$ , so that the final term of Equation 7 is 0. After  $M-2$  of these iterative steps, the library contains the optimal encoding scheme for all intervals  $[0, x]$  and  $M$  code words and, therefore, the optimal encoding scheme for the entire interval  $[0, 1]$ . At the iterative step of the algorithm, the procedure of inserting an additional cutpoint is performed not only for the entire odor range  $[0, 1]$ , but also for all subranges  $[0, x]$ . This is what enables a globally optimal solution to be obtained at the final step via a “look-back,” as diagrammed in Figure 3 (black arrows).

The algorithm is highly efficient compared with a brute-force approach. The number of computations scales with  $N^2 M$ , since there are  $M-2$  iterations, and the library has approximately  $N^2/2$  elements. This compares very favorably with a brute-force search of all possible encodings, which would require on the order of  $M^N$  computations.

Finally, we mention some points concerning generality of the approach. First, the algorithm applies not only to maximizing the mutual information (Eqs. 2 or 5), but also to a quantity (to be used below) that includes a penalty for codes that are complex, as follows:

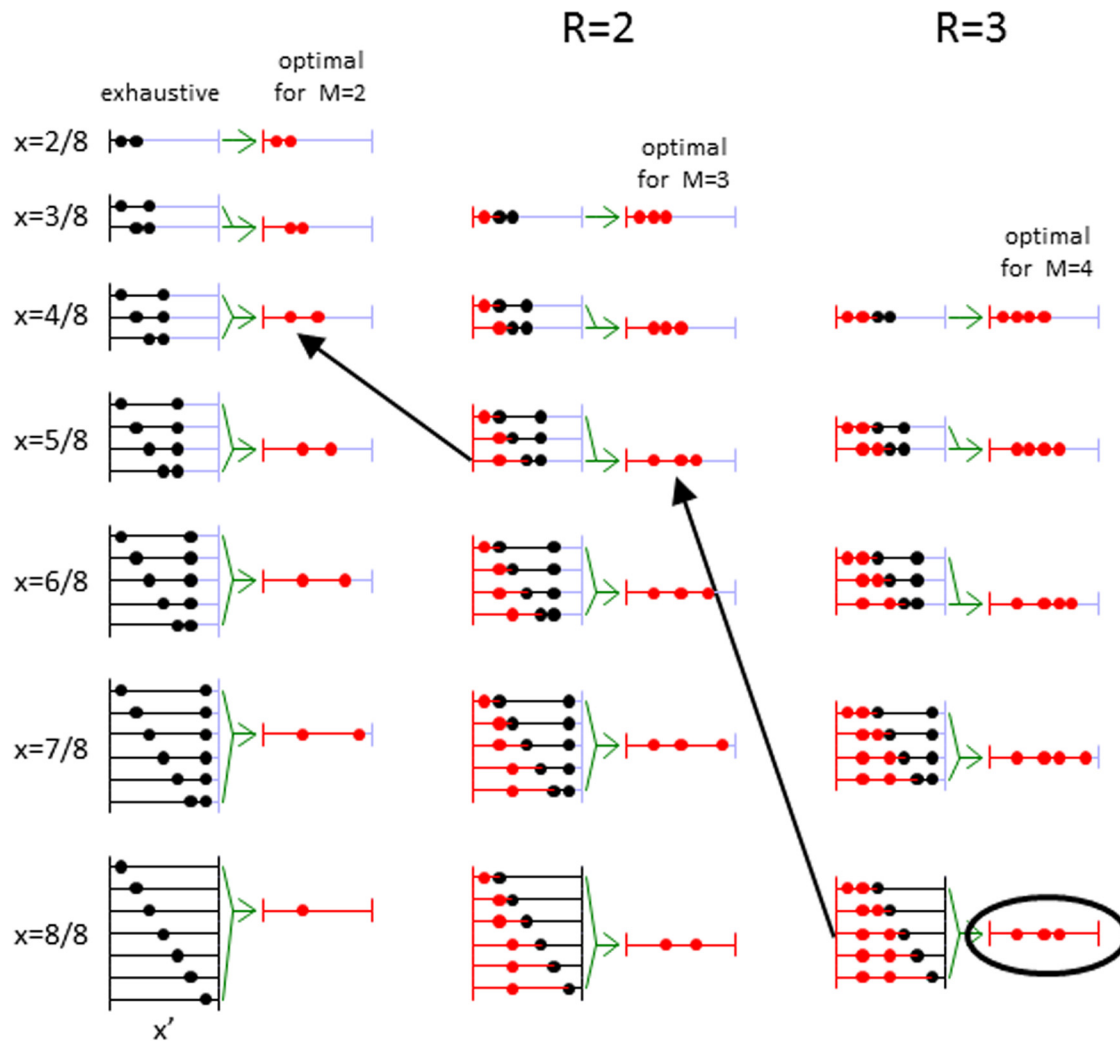
$$I_w(\mathbf{L}, \mathbf{M}) = (1-w)I_w(\mathbf{L}, \mathbf{M}) - wH(\mathbf{M}) \quad (8)$$

For  $w = 0$ ,  $I_w = I$ ; for  $w > 0$ , the entropy of the code word distribution  $H(\mathbf{M})$  acts as a penalty. The algorithm holds for Equation 8 since this quantity is also a linear combination of expressions in the form of Equation 6, and therefore also satisfies the chain rule.

We also note that the chain rule can be applied more generally, by partitioning the number of code words  $M$  into any two subsets of sizes  $M_u$  and  $M_v$ , for which  $M_u + M_v = M$ . This leads to a variant of the algorithm that require fewer iterative steps (approximately  $\log_2 M$  rather than  $M-2$ ) (for further analyses of this issue, see Osorio-Hernández et al., 2009; Clift, 2011). However, this benefit is outweighed by larger library size required (on the order of  $N^3$  rather than  $N^2$ ). We mention this variant architecture since it provides a starting point for further generalizations of the algorithm, but in the present context, it is less efficient and yields precisely the same results. Specifically, in our implementation and hardware (Intel i7 8-core processor), this variant was 20–100 times slower than the original.

*Finite sample size effects and statistics.* As is well known, empirical estimates of entropy and information are subject to bias due to finite sample size (Miller, 1955; Carlton, 1969; Treves and Panzeri, 1995; Strong et al., 1998). For the “plug-in” estimators used here, this bias, asymptotically in the number of samples, depends only on the number of samples, the number of input classes (here,  $L$ ), and the number of encoded code words (here,  $M$ ) (Miller, 1955; Carlton, 1969; Treves and Panzeri, 1995). Thus, in the asymptotic range, an optimization based on the “plug-in” estimator is equivalent to an optimization that takes bias correction into account.

The asymptotic range applies when the total number of observations is large compared with the product of the number of locations  $L$  and the number of encoded code words  $M$ . In the worst-case scenario here,  $L = 49$  and  $M = 64$ , so  $LM = 3136$ . The number of observations ( $T = 18,000$  to  $T = 36,000$  time points at  $L$  locations) is 882,000 to 1,764,000. Thus, this ratio is 281 for the shorter datasets and 562 for the longer ones.



**Figure 3.** Dynamic programming strategy for determining the partition of a range  $[0, 1]$  into code words that maximizes transmitted information, here illustrated for a resolution of  $N = 8$  and partitions into  $M = 4$  segments. The initial step (left column) creates a library of all intervals  $[0, x]$ , and all cutpoints  $x'$  within those intervals. For each interval  $[0, x]$ , the algorithm determines the value of  $x'$  that maximizes mutual information (green brackets and arrows). This yields a library (second column) containing the optimal partition of each subinterval  $[0, x]$  into 2 code words. Black represents the portion of the interval being analyzed. Red represents partitions that have been determined to be optimal. The first iterative step (columns labeled  $R = 2$ ) creates a library containing the optimal partitions of all intervals  $[0, x]$  into  $M = 3$  code words. It does this by subdividing each interval  $[0, x]$  into  $[0, x']$  and  $[x', x]$ , and using the previous library for the partitions of the  $[0, x']$  interval into 2 code words (red). The partition with the value of  $x'$  that maximizes information is then chosen for the new library (green brackets and arrows). Similarly, the next iterative step (columns labeled  $R = 3$ ) creates a library containing optimal partitions of all intervals  $[0, x]$  into  $M = 4$  code words, by subdividing each interval  $[0, x]$  into  $[0, x']$  and  $[x', x]$  using the previous library for the  $[0, x']$  interval. At the final stage, the optimal partition of the full interval  $[0, 1]$  (circled) can be traced back to optimal partitions determined at prior stages (large black arrows).

We empirically verified that the asymptotic range applied by analyzing full datasets, and also analyzing subsets of size  $T/2$  and  $T/4$  for the 2-dimensional grids, and we extrapolated values of  $I$  or  $I_w$  to infinite data by the method of Strong et al. (1998). These values, which are presented here, were only minimally different from the raw values determined with the full dataset, as would be expected from the large excess of data samples over the number of code words.

All calculations were performed in parallel at five sets of sample points: one set of points centered at the grid location, and four sets of sample points jittered diagonally by  $\pm 3$  pixels (2.2 mm) along each axis. Figure 4 shows results from the central point only; subsequent figures show an average of the results at the five jittered positions.

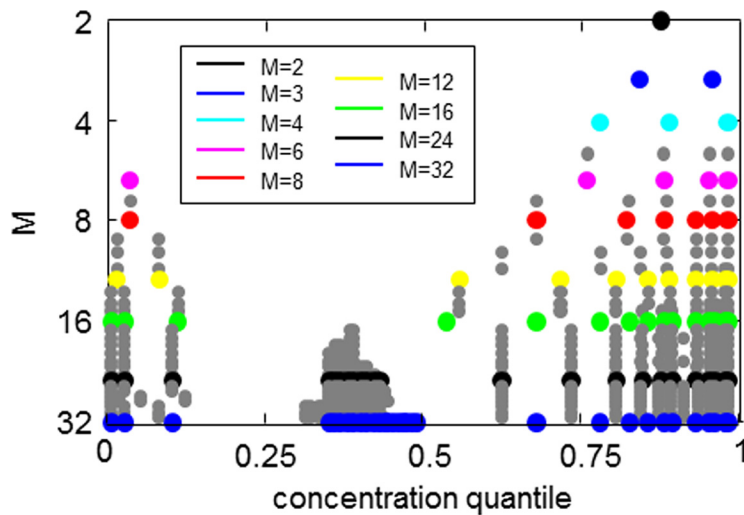
## Results

When an animal attempts to navigate toward the source of an odor plume, it makes navigation decisions based in part on sampling odor concentration. However, as with any sensory system, the resources available for signaling odor concentration are limited. Here, we focus on how the odor concentration in a single

plume sample can be encoded to maximize the amount of information transmitted about location relative to the source of the plume.

To address this question, we use the strategy shown in Figure 1. We begin with physical measurements of odor concentrations at a grid of points within a dynamic plume (Fig. 1A). Because the plume is dynamic, each point within the grid yields a probability distribution of measurements (Fig. 1B). Because these distributions vary across sampling points, a single sample provides a probabilistic cue to location. The amount of information about location is limited by the overlap of these distributions, and also by the resolution of the odor measurement.

To formalize the effects of limited resolution, we partition the range of odor concentration into a set of intervals, assuming that each interval is signaled by distinguishable neural “code words,” while two concentrations within a single interval cannot be resolved (Fig. 1C, left panels). For each such partitioning, we com-



**Figure 4.** Optimal segmentations for the 10 cm/s unbounded environment, full ( $7 \times 7$ ) grid. The positions of the cutpoints are shown for the number of code words ( $M$ ) ranging from 2 to 32, with a subset of values for  $M$  (colored points) highlighted at various levels on the ordinate. For each value of  $M$ , there are  $M-1$  cutpoints, which separate the full range of concentration into the segments corresponding to the code words.

pute the information transmitted about location, making optimal (Bayesian) use of the differences in the distributions of odor concentrations at each location (Fig. 1C, right panels). To remain agnostic about how this information is used, we quantify transmitted information by the Shannon mutual information measure (Shannon, 1948; Cover and Thomas, 1991).

The critical aspect of the analysis is to determine, for a fixed number of code words, how to segment the odor concentration range so that mutual information is maximized. That is, given a limit on the resources available for coding (the number of code words), how should these resources be deployed? To determine this optimal segmentation, we use a novel algorithm described in Materials and Methods, which solves this combinatorial optimization exactly. This analysis is then applied to physical measurements of odor concentrations in five olfactory environments, and nine grids of sample locations. Each sampling grid is merely a specific way to compare the distribution of odor concentrations across space; it does not imply a navigation strategy. As mentioned in Materials and Methods, our analysis considers each odor sample independently, and neglects correlations across time or space.

### Example dataset and sampling grid

We first present the analysis in detail for one environment and grid (10 cm/s unbounded, “full” grid) and then summarize the results across the environments and sampling grids.

Figure 4 shows the optimal segmentations of the odor concentration range as the number of available code words,  $M$ , increases from 2 (top of figure) to 32 (bottom of figure). The abscissa represents odor concentration as quantiles, so that the length of an interval (i.e., the distance between adjacent cutpoints) is proportional to the probability that the odor concentration is within that interval; and hence, the probability that the corresponding code word is used. In the simplest case ( $M = 2$ ), the cutpoint between the code words is approximately at the 85th percentile of odor concentration. That is, if only two alternatives (one bit) are available to signal odor concentration (either “low” or “high”), the optimal solution is to use the “high” code word only rarely, reserving it to the infrequent instances in which odor concentration is very high. This is in contrast to a coding strategy that

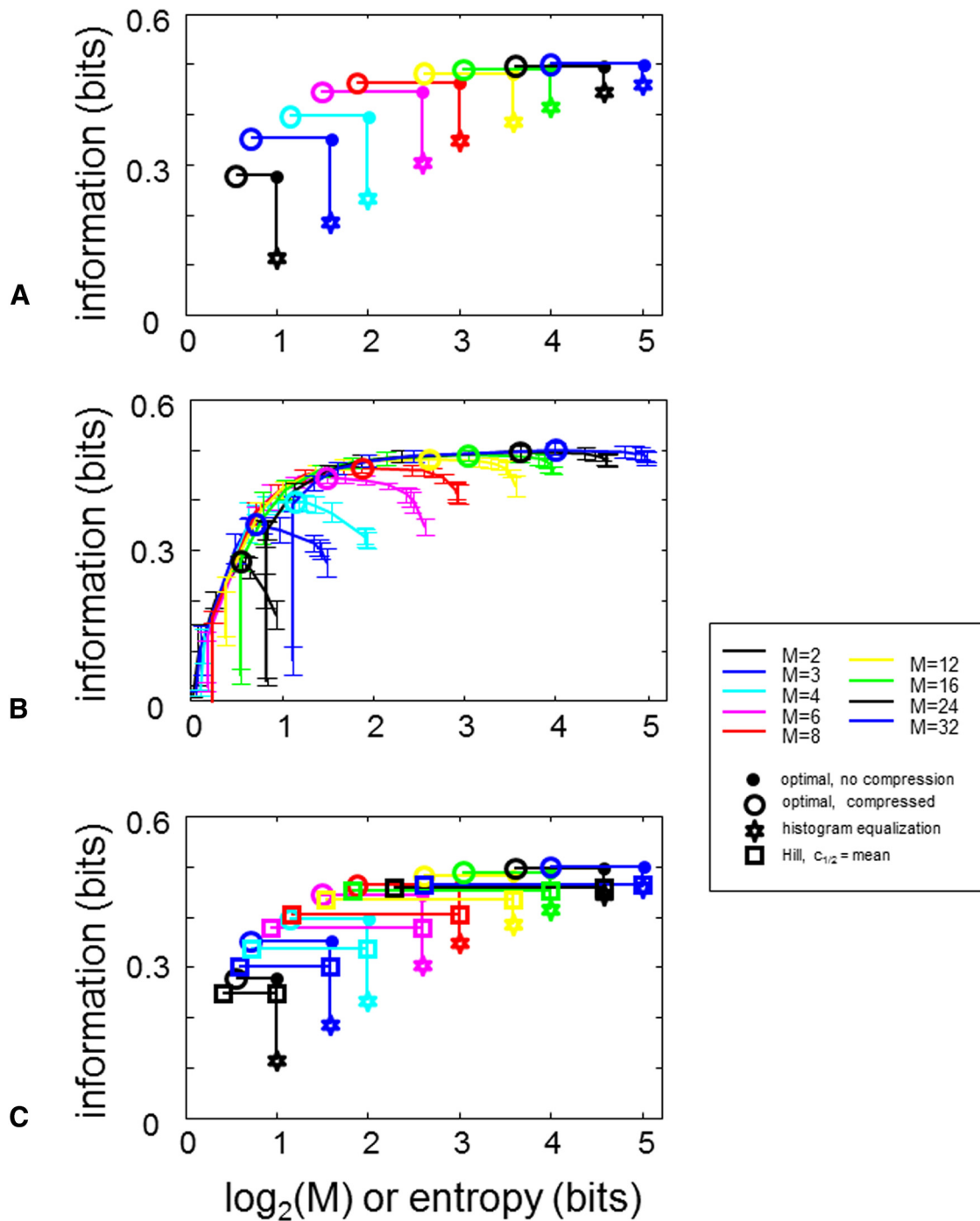
would be most effective in conveying information about odor concentration per se. For that purpose, the HE strategy is optimal (Laughlin, 1981): the cutpoint would be set at the median (i.e., the 50th percentile), so that the two code words “low” and “high” would be used equally often. This finding corroborates the result of Boie et al. (2018, their Fig. 4), in which the optimal binarizing cutpoint for conveying location information was determined by exhaustive search.

As the resources available to transmit odor concentration information increase, the range of odor concentration is subdivided into more and more code words, allowing for increasing resolution of odor concentration, but this increased resolution remains disproportionately focused on the high end of the concentration range. When three code words are available ( $M = 3$ ), the two cutpoints are positioned at approximately the 80th percentile and the

95th percentile. Consequently, the “medium” and “high” words, in aggregate, are only used  $\sim 20\%$  of the time, further contrasting with the HE strategy in which the cutpoints would be at the 33rd and 66th percentile. This exclusive emphasis on high concentrations persists until  $M = 6$ . At this point, cutpoints begin to appear at the low end of the concentration range as well, although the distribution of cutpoints remains heavily biased toward the high end of the range. In contrast, the HE strategy would cover the percentile range with intervals of equal size.

Next, we determine how effective these codes are in transmitting information about location (i.e., the mutual information between location and code word; Eq. 2). This is shown by Figure 5A (solid symbols), as a function of the number of bits in the code,  $\log_2 M$ . As anticipated from the findings of Boie et al. (2018), the amount of information transmitted saturates rapidly as a function of the number of code words available. Specifically, for  $M = 8$  (three bits),  $\sim 90\%$  of the maximum has been reached. An HE code eventually achieves the same level but requires a greater number of code words to do so (Fig. 5A, hexagrams).

The functional distinction between the optimal coding scheme and the HE scheme has another aspect, in addition to the gap between the optimal and HE information values shown in Figure 5A. This aspect is manifest when we consider how the code words themselves might be transmitted. As implied by the abscissa, it is always possible to specify one of  $M$  code words with  $\log_2 M$  bits, so  $\log_2 M$  is an upper bound on the number of bits needed to transmit a single code word. In the HE code, all code words are equally likely, so this bound cannot be improved on. However, in the optimal codes, some code words are much more common than others, which is a form of partial redundancy. As a consequence, typical sequences of code words can be reformatted (e.g., by downstream neural processing) so that fewer than  $\log_2 M$  bits per code word, on average, are required for transmission. If we neglect any temporal correlations in the sequence of code words, the limit of compression is given by the entropy of the code word distribution (Cover and Thomas, 1991). For the HE code, the entropy of the code word distribution is  $\log_2 M$  because each code word is equally likely; for the optimal codes, because the frequencies of the code words are typically unequal, it is lower.



**Figure 5.** *A*, Information transmitted about location by codes that are optimized for this purpose (filled and open round symbols), and HE (hexagrams), for the 10 cm/s unbounded environment and the full grid (as in Fig. 4). For filled symbols and hexagrams, the abscissa represents  $\log_2 M$ , where  $M$  is the number of code words. For open symbols, the abscissa represents the entropy of the code word distribution. (For the HE code, entropy is equal to  $\log_2 M$ ; for optimal codes, it is smaller.) Solid lines connect points with the same number of code words. *B*, For each number of code words, the locus of codes that optimize a tradeoff between information about location and compressibility. Each trajectory corresponds to a single value of  $M$ , following the color code of *A*. The left end of each trajectory corresponds to a maximally positive weighting for compressibility. Circle represents a weighting of 0 (matching the open symbols in *A*), and the right end of each trajectory corresponds to a large negative weighting for compressibility. *C*, Same as in *A*, but also including the performance of codes in which the Hill nonlinearity (Eq. 10) is followed by uniform segmentation of the  $[0, 1]$  range of bound fractions into  $M$  code words (open squares). For each value of  $M$ , the positions of the two squares along the abscissa indicate  $\log_2 M$  and the entropy of the code word distribution. Here and in subsequent figures, calculations are performed at the center of each grid point, and also four sets of sample points jittered diagonally by pixels (2.2 mm) along each axis. Mean values are shown in all panels. *B*, Error bars indicate 1 SD about the mean.

Figure 5A (open symbols) shows the implications of this compression for optimal codes. Because the optimal codes can be compressed but the HE code cannot, the difference between optimal codes and HE increase when compression is taken into

account: for the optimal codes, the number of bits required to transmit each code word is smaller than for HE.

This observation that performance of the “optimal” code was further improved when compression was taken into account mo-



tivated a broadening of our evaluation of possible coding schemes to explicitly include the value of compression (Tishby et al., 1999). To take compression into account, we identified coding schemes that maximized a function that included a penalty for the entropy of the code word distribution (Eq. 8). This penalty could take on a range of weights: for a weight  $w = 0$ , the analysis reduces to that of Figure 5A, in which compressibility was not taken into account; larger weights place progressively more emphasis on compressibility. The dynamic programming method described in Materials and Methods extends to this scenario as well, enabling identification of the optimal code for all tradeoffs between information and compressibility. For further details, see Materials and Methods, Equation 8, and surrounding text.

Figure 5B shows the results of the analysis. For each number  $M$  of code words, this procedure yields a family of codes that maximizes the various tradeoffs between information and compressibility. These codes range from highly compressible codes that provide very little information ( $w \rightarrow 1$ ), to codes that provide maximal information without taking compression into account at all ( $w = 0$ ). For values  $w < 0$ , the “optimal” codes give compressibility a negative weight, and tend toward HE, the least compressible code. Together, the envelope of these curves represents the codes that extremize the tradeoffs between information and compressibility. The codes identified in the original analysis (open circles) lie on this envelope, indicating that they capture the tradeoffs between information and compressibility.

### Coding based on the Hill nonlinearity

Finally, we consider a coding scheme motivated by a biological consideration, rather than an information-theoretic one. Because the information available to the organism is limited by what is taken in at the periphery (the “Data Processing Theorem”) (Cover and Thomas, 1991), it is natural to focus on the initial stage of olfactory transduction, which consists of an odorant ligand binding to a receptor (Nagel and Wilson, 2011; Cao et al., 2016; Gorur-Shandilya et al., 2017). The relationship between the concentration of the free ligand and fraction of receptors that have bound a ligand is nonlinear and saturating because the number of receptors is limited. For simple receptor binding and dissociation,



the equilibrium relationship is straightforward to determine. As is well known, under these circumstances, the bound fraction  $f$  corresponding to an odorant concentration  $[X]$  is given by the Hill equation with exponent 1 (Keener and Sneyd, 2010):

$$f([X]) = \frac{[X]}{c_{1/2} + [X]} \quad (10)$$

where  $c_{1/2}$ , the semisaturation constant, is given by  $c_{1/2} = k_D/k_B$ , the ratio of the dissociation rate constant (the reverse reaction in Eq. 9) and binding rate constant (the forward reaction in Eq. 9). As is also well known, the relationship (Eq. 10) also corresponds to the classic Michaelis–Menten kinetics (Michaelis et al., 2011) and to the Naka–Rushton adaptation law (Naka and Rushton, 1966).

We now consider a coding scheme in which, as in olfactory transduction, the nonlinearity (Eq. 10) is the first step, and its output is then encoded linearly. We take the semisaturation constant  $c_{1/2}$  equal to the mean odor concentration at the grid points, as this has been suggested as the target of sensory adaptation (Nagel et al., 2015; Cao et al., 2016; Gorur-Shandilya et al., 2017). To represent linear encoding of the results of the transduction

process, we posit that the bound fraction  $f$  is then encoded into  $M$  code words that subdivide the range equally. That is, for  $M = 2$ , values of  $f$  between 0 and  $1/2$  are encoded into one code word, and values of  $f$  between  $1/2$  and 1 are encoded into the other word. For a general  $M$ , values of  $f$  from  $k/M$  up to  $(k + 1)/M$  are encoded into the  $k$ th code word.

Figure 5C shows the performance of this code (square symbols), both without compression taken into account (rightmost symbol of each color pair) and with compression taken into account (leftmost symbol of each color pair). Performance is very similar to that of the optimal code identified from information-theoretic principles (filled circles represent without compression; open circles represent with compression).

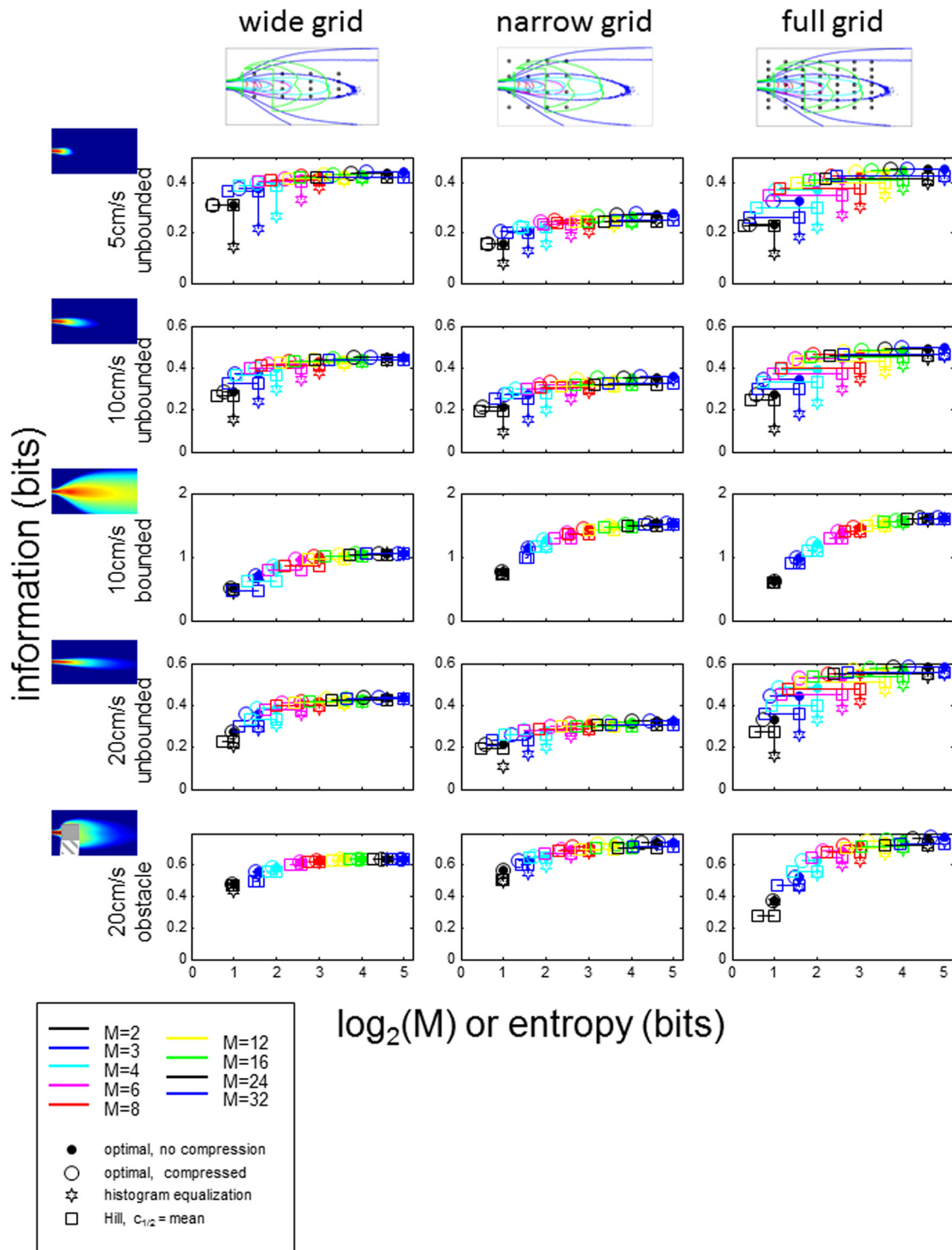
Like the information-theoretic optimal codes, the code determined by the Hill nonlinearity allocates a disproportionate fraction of code words to the rare high concentrations, compared with the HE code. In the Hill code, half of the code words are devoted to odor concentrations that are above  $c_{1/2}$  because that is the concentration at which the Hill nonlinearity (Eq. 10) reaches a value of 0.5. Since we have set  $c_{1/2}$  to be equal to the mean concentration—which is much higher than the median concentration in typical odor environments—these concentrations, and hence these code words, occur with probability much less than 0.5. Put another way, although the Hill nonlinearity is a saturating one, setting  $c_{1/2}$  at the mean concentration yields a nonlinearity that saturates much more gently than the nonlinearity that produces HE.

### Summary across datasets and grids

Figure 6 summarizes the above analysis across five odor environments and three sampling grids, and shows that the above observations hold across these scenarios. (The scenario detailed above is in the second row, third column.) In all cases, the Hill nonlinearity, with  $c_{1/2}$  set equal to the mean odor concentration across the sampling grid, yields a code that closely approximates the code that is optimized to transmit information about location. In most cases, the performance of these codes is substantially better than HE. The exception is the 10 cm/s bounded environment, in which all codes perform approximately equally well. These findings also hold for one-dimensional sampling grids, oriented either across or along the direction of bulk flow. Figure 7 shows results from the three example grids in the two environments with a 10 cm/s flow rate. In the unbounded environment (top row), there was a substantial advantage for the optimal code over the HE code for all sampling grids; in the bounded environment, the advantage was only present for the transverse grid placed closest to the source. Behavior for the other unbounded environments (5, 20, and 20 cm/s with an obstacle) was similar to that of the 10 cm/s environment. In all cases, the Hill-based code closely approximates optimal behavior.

In both Figures 6 and 7, the advantage of a code optimized for location is markedly diminished in the 10 cm/s bounded-flow environment. We speculate that the reason that the HE code is almost as effective as a code optimized for location is that this environment is the most diffusive, as the plume is near a boundary. Intuitively, in a diffusive environment, in which concentrations do not fluctuate over time, there is little difference between a code that is useful for reconstructing the plume (for which HE is optimized) and a code that is useful for navigating.

Figure 8, which shows the probability distribution of odor concentrations at a grid of points in the two 10 cm/s domains, amplifies this idea. In the diffusive environment (B), the distributions at different sample locations have markedly different modes, so these modes (i.e., the typical concentration at each



**Figure 6.** Performance of optimal codes for location in the five environments (rows), sampled at three two-dimensional grids (shown in column headers). Environments are illustrated by the insets at the left of each row showing heatmaps of average concentration (as in Fig. 2). Plotting conventions as in Figure 5C, which is reproduced here as the rightmost panel of the second row.

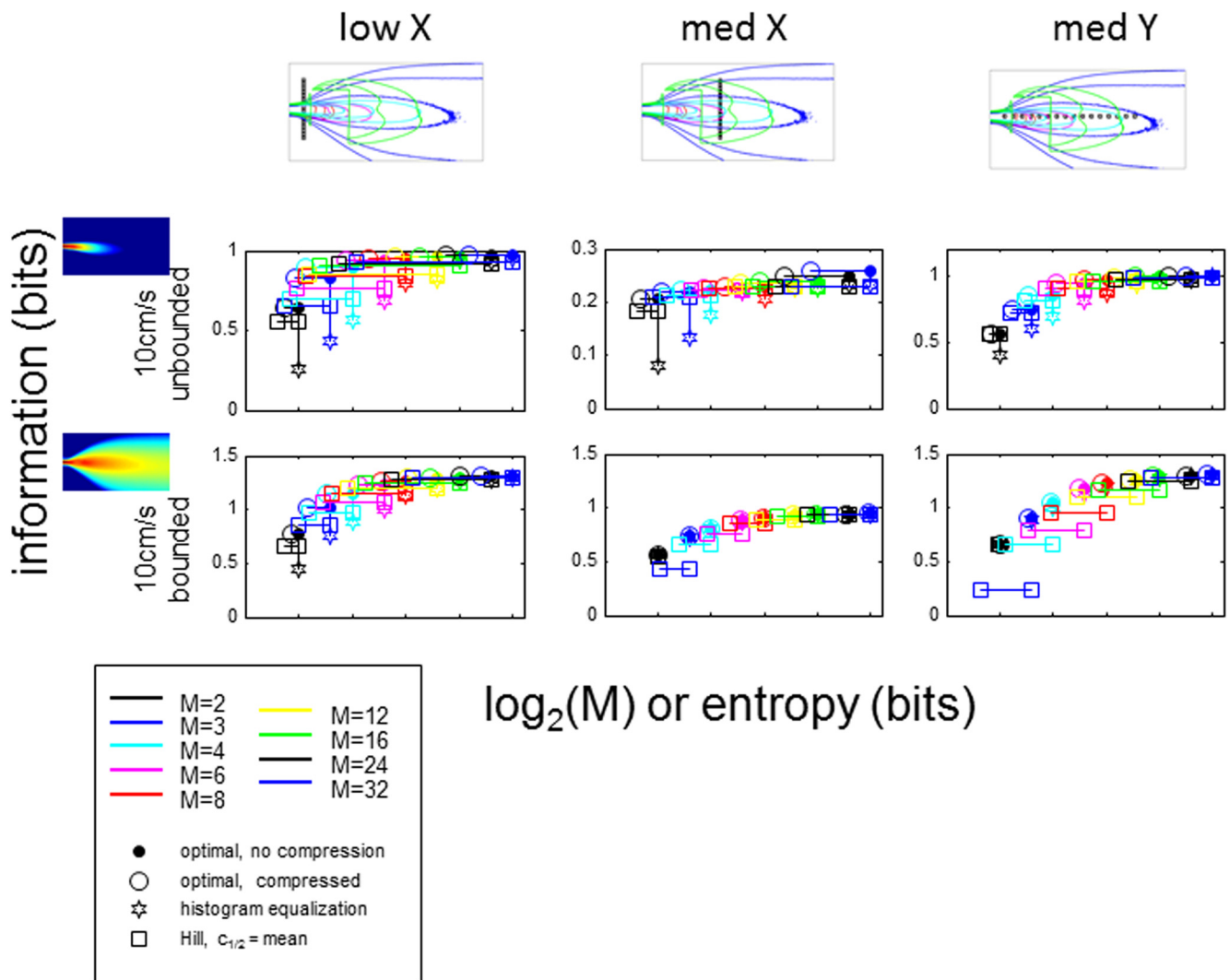
point) can be used as a cue to location. In contrast, in the turbulent environment (A), all of the modes are close to 0, so the typical odor concentration provides very little information. Instead, the main differences between the probability distributions are due to the concentrations that occur only rarely. That is, consistent with our findings above, location is cued primarily by the concentration values that have low probability.

However, one cannot conclude that the advantages of a code tuned for navigation are merely a matter of the degree of turbu-

lence; the advantage is smaller in the 20 cm/s environment with an obstacle (Fig. 6, bottom row) than in the unbounded environments.

**Role of the Hill constant**

Figures 6 and 7 show that the Hill nonlinearity, followed by linear coding, is near-optimal for transmitting information about location in a plume, across a range of odor environments. This result is based on setting the Hill semisaturation constant  $c_{1/2}$  equal to



**Figure 7.** Performance of optimal codes for location in two of the five environments, sampled at three one-dimensional grids (shown in column headers). Plotting conventions as in Figure 5C, but ordinate scales differ in all columns.

the mean odor concentration across the sampling grid. Because a real navigator cannot determine this value (it can only sample one location at a time), it is natural to ask whether this near-optimality is sensitive to a precise match of  $c_{1/2}$  to this mean value.

Figure 9 addresses this question for the full two-dimensional grid in the 10 cm/s unbounded environment. We measured the performance of codes in which the Hill constant differed from the mean odor concentration by a factor ranging from 0.125 to 8. In all cases, a mismatch by a factor of 2 (either half the grid mean or twice the grid mean) yields performance that is close to that setting  $c_{1/2}$  precisely equal to the grid mean. Performance decreases when the discrepancy is greater, but even when the discrepancy is a factor of 8, performance is better than that of HE. Similar results were found for the other environments and grids.

This finding implies a linkage between source concentration, plume statistics, and receptor sensitivities that yield near-optimum coding. For the plumes and sampling grids studied, the grid mean was typically 1%–10% of the source concentration (minimum: 0.45%; maximum: 21.3%; geometric mean: 2.4%). Receptor sensitivities vary widely, with typical Hill constants in the range of  $10^{-3}$  to  $10^{-6}$  (expressed as a volume dilution) but also extending lower (Si et al., 2019). Thus, the typical range of

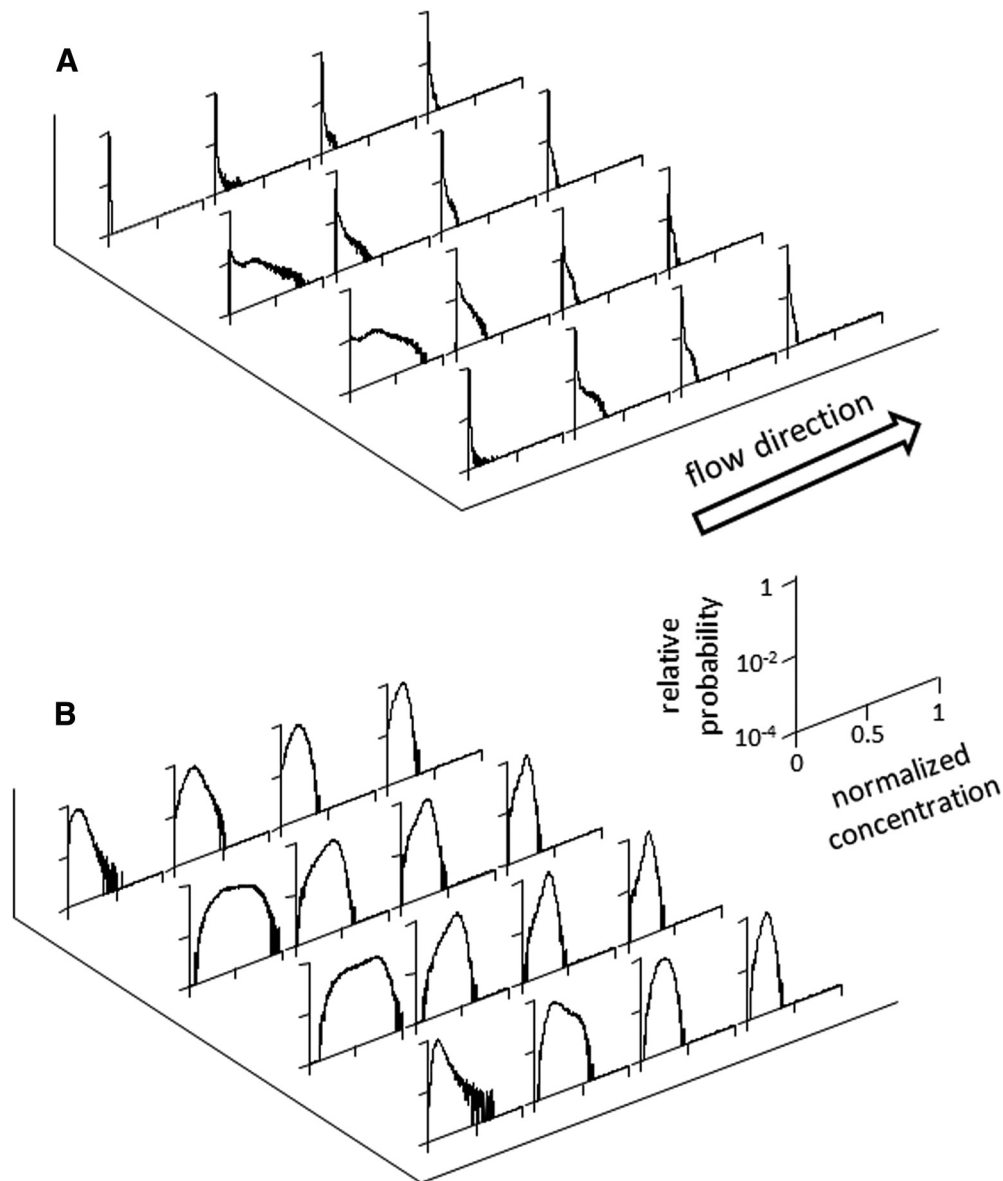
receptor sensitivities will allow for near-optimal coding in plumes whose source concentrations are  $10^{-1}$  to  $10^{-5}$  or below.

Finally, Figure 9 provides some insight as to the way in which coding performance changes when  $c_{1/2}$  departs from the grid mean. For values that are substantially smaller than the grid mean, the nonlinearity saturates rapidly, and behavior tends toward that of HE (Fig. 9, yellow points). For values that are substantially larger than the grid mean (Fig. 9, blue points), saturation becomes negligible, but only a small fraction of the code words are used because the bound fraction remains small. Thus, there is a ceiling to the effective amount of information carried because the effective number of code words is small.

## Discussion

### Efficient coding with a goal

Normative theories, approaches that identify the most efficient set of computations to accomplish a task and then compare these computations with those that are actually performed, are central to understanding principles of sensory processing (Barlow, 1961; Laughlin, 1981; Srinivasan et al., 1982; Olshausen and Field, 1996; Karklin and Lewicki, 2009; Tkacik et al., 2010; Hermundstad et al., 2014). A common feature of these theories is a focus on



**Figure 8.** Probability distributions of concentration at a grid of points (the wide grid shown at the top of the first column of Fig. 6) for the 10 cm/s unbounded environment (**A**) and the 10 cm/s bounded environment (**B**). Abscissa represents normalized concentration (concentration relative to concentration at the source) on a linear scale. Ordinate is logarithmic and spans a factor of  $10^4$ .

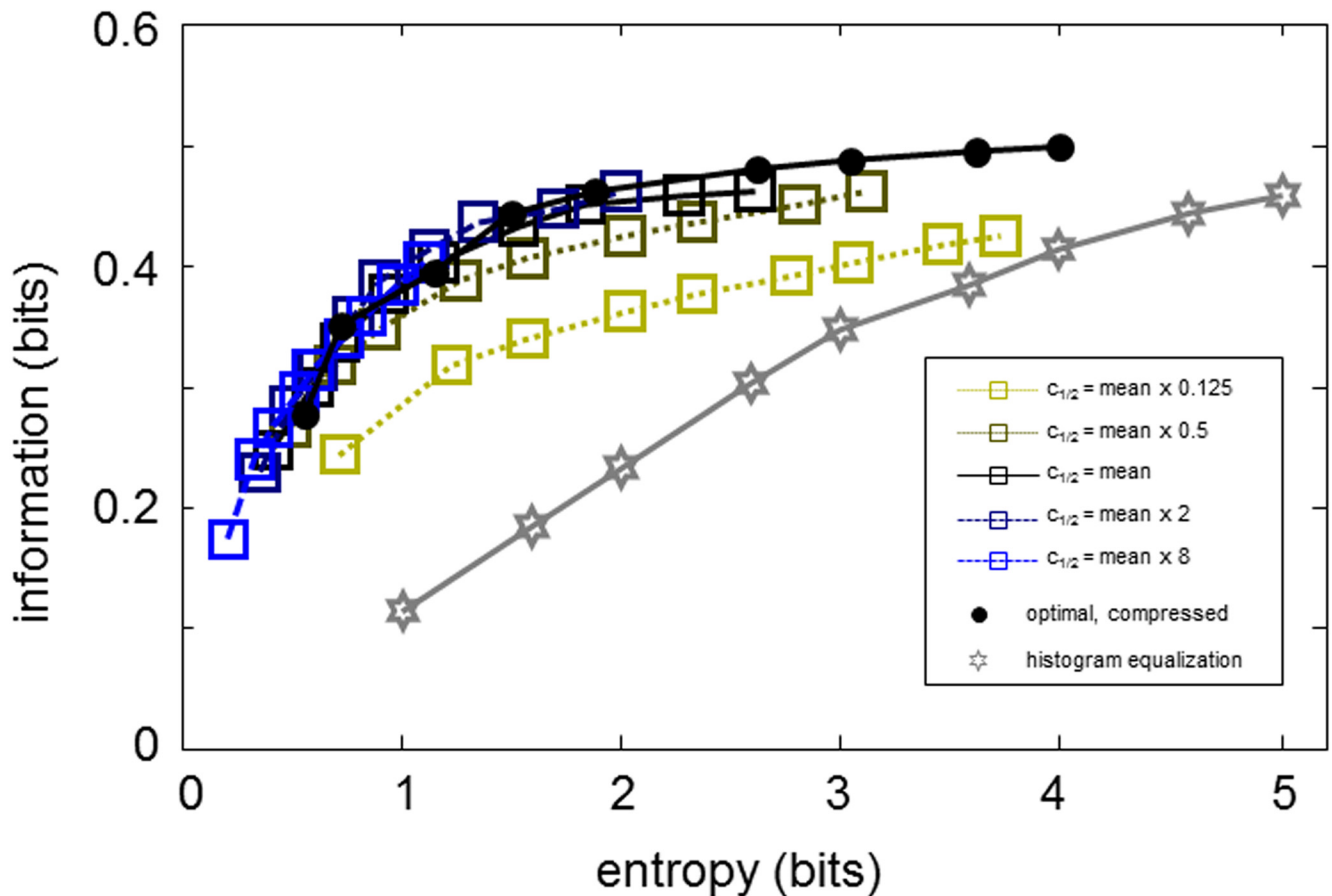
representing the sensory input, or estimating a simple statistic of its distribution (Fairhall et al., 2001; Wark et al., 2009; Mlynarski and Hermundstad, 2018), rather than the needs of a particular behavior. That is, the measure of efficiency is stimulus-centered: reconstruction of the input (Laughlin, 1981; Olshausen and Field, 1996), or discriminating among alternatives (Karklin and Lewicki, 2009; Hermundstad et al., 2014; Mlynarski and Hermundstad, 2018).

Olfactory navigation provides an opportunity to extend the normative approach to a context in which the driver of the behavior (source location) is linked to the sensory input in a complex and stochastic fashion. Creating an internal representation of the plume, estimating its statistics, or discriminating among odor concentrations is only helpful to the extent that it reduces uncertainty about source location.

Recognizing that any sensory system has limited resources for coding (i.e., a limited range and resolution), we determined how these resources are best deployed to convey information about

the navigator's location relative to the source. We found that the optimal coding strategy differed from one that was optimal for creating an internal representation, devoting greater resources to higher odor concentrations. Moreover, a mechanism known to be present at the first stage of olfactory processing, receptor binding (Kaissling, 1986; Nagel and Wilson, 2011; Gorur-Shandilya et al., 2017), constitutes a simple nonlinearity that closely approximates the identified optimal behavior. The key feature that underlies this result is that the Hill nonlinearity compresses the range of odor concentrations, but does so in a gentle fashion, so that high concentrations can still be distinguished. Other monotonic functions with similar shapes, including Hill functions with exponents modestly  $>1$ , would likely achieve near-optimal behavior as well.

Because any single sample of odor concentration provides only partial information about source location, successful olfactory navigation requires integration of many odor samples, and possibly cues from other modalities (Murlis et al., 1992; Cardé



**Figure 9.** Performance of codes based on the Hill nonlinearity with several choices for the semisaturation constant  $c_{1/2}$  in the 10 cm/s unbounded environment, sampled by the full grid. Black squares represent  $c_{1/2}$  set equal to the mean across grid points. Blue squares represent  $c_{1/2}$  larger than the mean. Yellow squares represent  $c_{1/2}$  smaller than the mean. For comparison, filled circles represent the optimal code for location taking compression into account (corresponding to Fig. 5C, open circles). Hexagrams represent the HE code (as in Fig. 5C). Abscissa represents entropy of code word distribution.

and Willis, 2008; Gaudry et al., 2012; Alvarez-Salvado et al., 2018). Such strategies may take several forms, ranging from construction of a cognitive map (Vergassola et al., 2007; Gagliardo, 2013; Grünbaum and Willis, 2015; Jacobs et al., 2015) to local decisions, which may depend on an internal state defined by recent history (Gaudry et al., 2012; van Breugel and Dickinson, 2014; Schulze et al., 2015; Kromer et al., 2018; Pang et al., 2018). Because our approach considers only single samples and hence ignores spatial or temporal correlations in the odor plume, it does not address ways in which encoding may be further tuned to, or interact with, an overall search strategy. Nevertheless, we expect that the findings here are broadly relevant because, in any search strategy, encoding an odor sample is an unavoidable first step.

This first step consists of representing an odor concentration by a neural signal. In Boie et al. (2018), we showed that precise representation of concentration has limited value in determining location. Beyond 16–32 coding levels (4–5 bits), additional resolution provided minimal additional information, but this single-sample ceiling could be overcome by devoting the bits to two samples in space, or multiple samples in time (Boie et al., 2018, their Fig. 3). That analysis considered the standard HE strategy, in which each code word is used equally often: a scheme that is optimal for stimulus reconstruction. Here, we show that the single-sample ceiling can be reached with only 4–8 coding levels (2–3 bits), by allocating more code words to the higher concentrations (Figs. 6, 7), especially in the more turbulent reg-

imens. However, “task-specific” coding does not change the amount of information about location; rather, it provides the same information in fewer bits (Boie et al., 2018). The advantage is that since fewer bits suffice for each sample, the same bandwidth can be redeployed to convey information from multiple samples in space and/or time.

We focused on the nonlinearity associated with transduction because this places a fundamental limit on the information that is available to the organism, but clearly, this is only the beginning of the computations that underlie olfactory navigation. Subsequent stages of processing add further transformations, and interactions between odorants, for example, divisive normalization at the convergence of olfactory receptor neurons onto projection neurons (Olsen et al., 2010). Because this transformation expands small signals and compresses large signals, the signals sent on to the brain will differ from a simple Hill transformation, and will also be modulated by signals from other receptors.

#### Dynamic adaptation and search

While the Hill nonlinearity (Eq. 10) comes close to an optimally efficient transformation of odor concentration to coding level, this behavior requires (Fig. 9) that the semisaturation constant,  $c_{1/2} = k_D/k_B$ , is close to the “grid mean” (i.e., the mean concentration of a specific odorant across sampling points of a particular odor environment). This in turn requires that it is set dynamically, rather than fixed by evolution or development. Adaptation

that sets  $c_{1/2}$  to the mean with a slow time constant is well recognized (Nagel et al., 2015; Cao et al., 2016; Gorur-Shandilya et al., 2017), but there is still the question of how this mean is determined. Each grid has its own mean concentration; and even with only one grid, the navigator can only be at one location at any given time. We speculate that this consideration provides a relationship between the dynamics of adaptation and the dynamics of the search strategy. Specifically, we suggest that olfactory navigation can be considered as a sequence of searches at progressively smaller spatial scales, in which the navigator visits a variety of locations (“grid points”) at one scale and then narrows the search. If the time spent searching at each scale is comparable to the time constant for adaptation, an approximate correspondence between  $c_{1/2}$  and the grid mean will be achieved. Only an approximate correspondence is needed for near-optimal behavior, as mismatches between  $c_{1/2}$  and the grid mean up to a factor of  $\sim 2$  have little effect (Fig. 9).

Values of  $c_{1/2}$  that are substantially lower than the mean produce a rapid saturation that is suboptimal for navigation. But at the same time, this more rapid saturation provides an approximation to the HE code that is optimal for transmitting information about concentration. Thus, dynamic adjustment of  $c_{1/2}$ , as well as a diversity of values of  $c_{1/2}$  across receptors, could tune the encoding process to diverse demands.

### Other domains

The present analytic approach applies not only to olfactory navigation, but to any domain requiring inference of location from a scalar variable. Such domains include chemotaxis in its many forms (e.g., Webre et al., 2003; Larsch et al., 2015; Schulze et al., 2015; Kromer et al., 2018), as well as navigation driven by temperature (Luo et al., 2010; Klein et al., 2015). As in olfactory navigation, the first processing stage involves a receptor. However, whether the nonlinearity of receptor binding plays a similarly important computational role must remain a speculation, as the benefit of the Hill nonlinearity depends on how the distributions of the scalar quantity vary across space. The pattern of variation found in olfactory plumes, where dynamics are dominated by turbulent fluid flow, may not be shared in scenarios in which flow is more laminar or diffusion is more prominent. It is also unclear whether the benefits of the Hill nonlinearity for independent receptor-binding sites (Eq. 10) extend to situations in which receptor binding is phenomenologically cooperative, as in the case of the TRP receptor (Zheng, 2013).

### Related information-theoretic approaches

The present approach bears similarity to the information-bottleneck (IB) (Tishby et al., 1999) and codebook (Dimitrov et al., 2003) methods. These are closely related approaches but have different goals: IB, like the present approach, seeks efficient ways of encoding a stimulus to convey information about an underlying variable, whereas the codebook method is a neural data analysis strategy that seeks categories within a stimulus set that best account for the neural response.

Both approaches differ from the present approach by their choice of cost function: they add a penalty when the mutual information between the sensory variable and the code word is high. Consequently, IB typically identifies encoding schemes that are nondeterministic. In contrast, here we add no penalty for reliable encoding. That is, we maximize the information conveyed about location, rather than minimize the information conveyed about concentration. Both approaches examine the tradeoff between coding fidelity and information transmitted

about location, albeit in different ways: IB does so explicitly via a penalty term; the present approach does so implicitly by considering code word sets of many sizes.

Related to these differences, algorithms to optimize encoding in the IB sense use simulated annealing, rather than the dynamic programming algorithm introduced here. If the annealing process could be continued down to a temperature of 0, the IB cost function would coincide with the present one, and the solutions will coincide. However, this strategy is intractable because of the explosion of bifurcations that would need to be followed (Dimitrov et al., 2003; Mumei and Gedeon, 2003). In contrast, the present algorithm finds the optimal coding scheme with a computational cost that grows quadratically with the resolution of the analysis.

In conclusion, we used olfactory navigation to study neural coding in the service of a biologically important task: identifying the relative location of the source of an odor plume. Across a broad range of conditions, we found that the nonlinearity associated with receptor binding yields an encoding that is nearly optimal for this purpose. Thus, receptor binding, along with its necessary mechanistic role in sensory processing, also plays a critical information-processing role.

### References

- Alvarez-Salvado E, Licata AM, Connor EG, McHugh MK, King BM, Stavropoulos N, Victor JD, Crimaldi JP, Nagel KI (2018) Elementary sensory-motor transformations underlying olfactory navigation in walking fruit-flies. *Elife* 7:e37815.
- Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: *Sensory communication*. Cambridge, MA: Massachusetts Institute of Technology.
- Boie SD, Connor EG, McHugh M, Nagel KI, Ermentrout GB, Crimaldi JP, Victor JD (2018) Information-theoretic analysis of realistic odor plumes: what cues are useful for determining location? *PLoS Comput Biol* 14:e1006275.
- Cao LH, Jing BY, Yang D, Zeng X, Shen Y, Tu Y, Luo DG (2016) Distinct signaling of *Drosophila* chemoreceptors in olfactory sensory neurons. *Proc Natl Acad Sci U S A* 113:E902–E911.
- Cardé RT, Willis MA (2008) Navigational strategies used by insects to find distant, wind-borne sources of odor. *J Chem Ecol* 34:854–866.
- Carlton AG (1969) On the bias of information estimates. *Psychol Bull* 71:108–109.
- Celani A, Villermaux E, Vergassola, M (2014) Odor landscapes in turbulent environments. *Phys Rev* 4:041015.
- Clift NM (2011) Calculating optimal addition chains. *Computing* 91:265–284.
- Connor EG, McHugh M, Crimaldi, JP (2018) Quantification of airborne odor plumes using planar laser-induced fluorescence. *Exp Fluids* 59:137.
- Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley.
- Dimitrov AG, Miller JP, Gedeon T, Aldworth Z, Parker AE (2003) Analysis of neural coding through quantization with an information-based distortion measure. *Network* 14:151–176.
- Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* 412:787–792.
- Gagliardo A (2013) Forty years of olfactory navigation in birds. *J Exp Biol* 216:2165–2171.
- Gaudry Q, Nagel KI, Wilson RI (2012) Smelling on the fly: sensory cues and strategies for olfactory navigation in *Drosophila*. *Curr Opin Neurobiol* 22:216–222.
- Gorur-Shandilya S, Demir M, Long J, Clark DA, Emonet T (2017) Olfactory receptor neurons use gain control and complementary kinetics to encode intermittent odorant stimuli. *Elife* 6:e27670.
- Grünbaum D, Willis MA (2015) Spatial memory-based behaviors for locating sources of odor plumes. *Mov Ecol* 3:11.
- Hermundstad AM, Briguglio JJ, Conte MM, Victor JD, Balasubramanian V, Tkacik G (2014) Variance predicts salience in central sensory processing. *Elife* 3:03722.

- Jacobs LF, Arter J, Cook A, Sulloway FJ (2015) Olfactory orientation and navigation in humans. *PLoS One* 10:e0129387.
- Kaissling KE (1986) Chemo-electrical transduction in insect olfactory receptors. *Annu Rev Neurosci* 9:121–145.
- Karklin Y, Lewicki MS (2009) Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457:83–86.
- Keener JP, Sneyd J (2010) *Mathematical physiology 1: Cellular physiology*. New York: Springer.
- Klein M, Afonso B, Vonner AJ, Hernandez-Nunez L, Berck M, Tabone CJ, Kane EA, Pieribone VA, Nitabach MN, Cardona A, Zlatic M, Sprecher SG, Gershow M, Garrity PA, Samuel AD (2015) Sensory determinants of behavioral dynamics in *Drosophila* thermotaxis. *Proc Natl Acad Sci U S A* 112:E220–E229.
- Kleinberg J, Tardos E (2006) *Algorithm design*. Boston: Pearson Addison Wesley.
- Kromer JA, Märcker S, Lange S, Baier C, Friedrich BM (2018) Decision making improves sperm chemotaxis in the presence of noise. *PLoS Comput Biol* 14:e1006109.
- Larsch J, Flavell SW, Liu Q, Gordus A, Albrecht DR, Bargmann CI (2015) A circuit for gradient climbing in *C. elegans* chemotaxis. *Cell Rep* 12:1748–1760.
- Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch C* 36:910–912.
- Luo L, Gershow M, Rosenzweig M, Kang K, Fang-Yen C, Garrity PA, Samuel AD (2010) Navigational decision making in *Drosophila* thermotaxis. *J Neurosci* 30:4261–4272.
- Michaelis L, Menten ML, Johnson KA, Goody RS (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* 50:8264–8269.
- Miller GA (1955) Note on the bias on information estimates. In: *Information theory in psychology: problems and methods II-B* (Quastler H, ed), pp 95–100. Glencoe, IL: Free.
- Mlynarski WF, Hermundstad AM (2018) Adaptive coding for dynamic sensory inference. *Elife* 7:e32055.
- Mumey B, Gedeon T (2003) Information quantization is NP-complete, NIPS 2003. Vancouver: Neural Information Processing Systems.
- Murlis J, Elkinton JS, Carde, RT (1992) Odor plumes and how insects use them. *Annu Rev Entomol* 37:505–532.
- Nagel KI, Wilson RI (2011) Biophysical mechanisms underlying olfactory receptor neuron dynamics. *Nat Neurosci* 14:208–216.
- Nagel KI, Hong EJ, Wilson RI (2015) Synaptic and circuit mechanisms promoting broadband transmission of olfactory stimulus dynamics. *Nat Neurosci* 18:56–65.
- Naka KI, Rushton WA (1966) S-potentials from luminosity units in the retina of fish (*Cyprinidae*). *J Physiol* 185:587–599.
- Olsen SR, Bhandawat V, Wilson RI (2010) Divisive normalization in olfactory population codes. *Neuron* 66:287–299.
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
- Osorio-Hernández L, Mezura-Montes E, Cruz-Cortés, N, Rodríguez-Henríquez, F (2009) A genetic algorithm with repair and local search mechanisms able to find minimal length addition chains for small exponents. Trondheim, Norway: IEEE Congress on Evolutionary Computation.
- Pang R, van Breugel F, Dickinson M, Riffell JA, Fairhall A (2018) History dependence in insect flight decisions during odor tracking. *PLoS Comput Biol* 14:e1005969.
- Pierce-Shimomura JT, Morse TM, Lockery SR (1999) The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *J Neurosci* 19:9557–9569.
- Schulze A, Gomez-Marin A, Rajendran VG, Lott G, Musy M, Ahammad P, Deogade A, Sharpe J, Riedl J, Jarriault D, Trautman ET, Werner C, Venkadesan M, Druckmann S, Jayaraman V, Louis M (2015) Dynamical feature extraction at the sensory periphery guides chemotaxis. *Elife* 4:06694.
- Shannon C (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423.
- Si G, Kanwal JK, Hu Y, Tabone CJ, Baron J, Berck M, Vignoud G, Samuel AD (2019) Structured odorant response patterns across a complete olfactory receptor neuron population. *Neuron*. Advance online publication. Retrieved January 16, 2019. doi: 10.1016/j.neuron.2018.12.030.
- Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci* 216:427–459.
- Strong SP, Koberle R, Ruyter van Steveninck, RR, Bialek, W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.
- Tadmor Y, Tolhurst DJ (2000) Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Res* 40:3145–3157.
- Tishby N, Pereira FC, Bialek, W (1999) The Information Bottleneck Method. 37th Allerton Conference on communication, control, and computing. Urbana-Champaign, IL: University of Illinois.
- Tkacik G, Prentice JS, Victor JD, Balasubramanian V (2010) Local statistics in natural scenes predict the saliency of synthetic textures. *Proc Natl Acad Sci U S A* 107:18149–18154.
- Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. *Neural Comput* 7:399–407.
- van Breugel F, Dickinson MH (2014) Plume-tracking behavior of flying *Drosophila* emerges from a set of distinct sensory-motor reflexes. *Curr Biol* 24:274–286.
- Vergassola M, Villermaux E, Shraiman BI (2007) 'Infotaxis' as a strategy for searching without gradients. *Nature* 445:406–409.
- Wark B, Fairhall A, Rieke F (2009) Timescales of inference in visual adaptation. *Neuron* 61:750–761.
- Webre DJ, Wolanin PM, Stock JB (2003) Bacterial chemotaxis. *Curr Biol* 13:R47–R49.
- Wen Q, Po MD, Hulme E, Chen S, Liu X, Kwok SW, Gershow M, Leifer AM, Butler V, Fang-Yen C, Kawano T, Schafer WR, Whitesides G, Wyart M, Chklovskii DB, Zhen M, Samuel AD (2012) Proprioceptive coupling within motor neurons drives *C. elegans* forward locomotion. *Neuron* 76:750–761.
- Zheng J (2013) Molecular mechanism of TRP channels. *Compr Physiol* 3:221–242.