Anticipatory Neural Activity Improves the Decoding Accuracy for Dynamic Head-Direction Signals

Johannes Zirkelbach, Martin Stemmler, and Andreas V.M. Herz

Bernstein Center for Computational Neuroscience Munich and Faculty of Biology, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

Insects and vertebrates harbor specific neurons that encode the animal's head direction (HD) and provide an internal compass for spatial navigation. Each HD cell fires most strongly in one preferred direction. As the animal turns its head, however, HD cells in rat anterodorsal thalamic nucleus (ADN) and other brain areas fire already before their preferred direction is reached, as if the neurons anticipated the future HD. This phenomenon has been explained at a mechanistic level, but a functional interpretation is still missing. To close this gap, we use a computational approach based on the movement statistics of male rats and a simple model for the neural responses within the ADN HD network. Network activity is read out using population vectors in a biologically plausible manner, so that only past spikes are taken into account. We find that anticipatory firing improves the representation of the present HD by reducing the motion-induced temporal bias inherent in causal decoding. The amount of anticipation observed in ADN enhances the precision of the HD compass read-out by up to 40%. More generally, our theoretical framework predicts that neural integration times not only reflect biophysical constraints, but also the statistics of behaviorally relevant stimuli; in particular, anticipatory tuning should be found wherever neurons encode sensory signals that change gradually in time.

Key words: anticipation; computational modeling; head-direction system; neural population activity; optimal decoding; rodent anterior dorsal nucleus

Significance Statement

Across different brain regions, populations of noisy neurons encode dynamically changing stimuli. Decoding a time-varying stimulus from the population response involves a trade-off: For short read-out times, stimulus estimates are unreliable as the number of stochastic spikes is small; for long read-outs, estimates are biased because they lag behind the true stimulus. We show that optimal decoding of temporally correlated stimuli not only relies on finding the right read-out time window but requires neurons to anticipate future stimulus values. We apply this general framework to the rodent head-direction system and show that the experimentally observed anticipation of future head directions can be explained at a quantitative level from the neuronal tuning properties, network size, and the animal's head-movement statistics.

Introduction

Spatial orientation requires information about one's current position and heading direction. In rodents, positional information is provided by place cells (O'Keefe and Dostrovsky, 1971), grid cells (Hafting et al., 2005), boundary vector and border cells (Barry et al., 2006; Solstad et al., 2008), and further spatially mod-

ulated neurons. As in other vertebrate (Finkelstein et al., 2015) and insect (Seelig and Jayaraman, 2015; Varga and Ritzmann, 2016) species, directional information is represented by head-direction (HD) cells across various brain areas (Taube et al., 1990a; Wiener and Taube, 2005; Taube, 2007).

HD cells integrate motor efference copies with vestibular, visual, and proprioceptive signals, and are maximally active when the animal's head points into a cell-specific preferred direction (Taube, 2007). These directions are uniformly distributed and rotate in close synchrony with rotated salient visual cues. During navigation in darkness, direction-specific firing is maintained, but the preferred directions drift over time (Taube et al., 1990b). Optogenetically interfering with the HD system impairs homing, demonstrating that the HD system plays a crucial role for spatial navigation (Butler et al., 2017).

Conceptually, one can arrange the preferred directions of HD cells on a ring that is anchored to the local environment (see Fig.

Received Oct. 8, 2018; revised Dec. 16, 2018; accepted Jan. 11, 2019.

Author contributions: J.Z. wrote the first draft of the paper; J.Z., M.S., and A.V.M.H. edited the paper; J.Z., M.S., and A.V.M.H. designed research; J.Z., M.S., and A.V.M.H. performed research; J.Z. analyzed data; J.Z., M.S., and A.V.M.H. wrote the paper.

This work was supported by the German Federal Ministry for Education and Research Grants 01GQ0440. We thank E.I. Moser and G. Buzsáki for making data from Sargolini et al. (2006) and Mizuseki et al. (2009), respectively, publicly available; and A. Loebel and A. Mathis for stimulating discussions.

The authors declare no competing financial interests.

J. Zirkelbach's present address: Max Planck Institute for the Science of Light, Staudtstraße 2, 91058 Erlangen, Germany.

Correspondence should be addressed to Andreas V.M. Herz at herz@bio.lmu.de.

<https://doi.org/10.1523/JNEUROSCI.2605-18.2019>

Copyright © 2019 the authors

Table 1. Properties of HD cells in the ADN

Parameter	Mean	Standard Deviation	Range
f_{max}	50 Hz	27 Hz	5–130 Hz
f_{bg}	2 Hz	2 Hz	0–10 Hz
σ	25°	5°	15°–35°
τ	25 ms	15 ms	–10 to 100 ms

Summary of HD cell parameters as reported in the ADN literature (Blair and Sharp, 1995; Goodridge and Taube, 1997; Blair et al., 1997, 1998; Stackman and Taube, 1998; Taube and Muller, 1998; Taube, 2007; Clark et al., 2009). f_{max} and f_{bg} , Peak and background firing rates, respectively; σ , tuning width; τ , anticipatory time interval. The concentration parameter κ of von Mises tuning curves is given by $\kappa = (\sigma \times \pi/180)^{-2}$ so that κ ranges from ~ 2.7 (for $\sigma = 35^\circ$) to ~ 14.6 (for $\sigma = 15^\circ$) and the mean σ value corresponds to $\kappa \approx 5.25$.

1). When the animal turns its head, the packet of active neurons is updated such that those cells are most active whose preferred directions match the animal's new HD. As the HD can be read out from the position of this activity packet along the ring, the HD system is often referred to as an internal or neural compass (Schultheiss and Redish, 2015). Unlike the needle of a physical compass, which points toward magnetic north, the preferred directions of HD cells are determined by local cues and vary from environment to environment.

How accurately can an animal determine its HD from this compass? Typical tuning widths of individual HD cells (see Figs. 1B, 2) range from 30° to 70° (see Table 1), too large for accurate navigation on the basis of single-cell responses. However, combining the information provided by all simultaneously active HD cells leads to a far more reliable estimate. To understand HD coding at the network level, we construct a computational model, read out its activity using population-vector decoding (Georgopoulos et al., 1983; Seung and Sompolinsky, 1993; Salinas and Abbott, 1994), and study the accuracy and bias of this neural compass. For concreteness, we focus on the anterodorsal thalamic nucleus (ADN), which is a core region of the HD system (Taube, 2007) and contains the highest fraction of HD cells (Taube and Bassett, 2003).

During foraging, the animal's HD becomes a time-varying quantity. In this case, causal decoding, for which only spikes from the past are used, biases the population vector, as we will show. To counteract this temporal bias, a decoder can reduce the read-out time, yet shorter read-out times result in higher variability.

Interestingly, HD cells in the ADN fire in an anticipatory manner: during head turns, their preferred directions shift such that the cells fire earlier than without movement, as though they encoded the future HD (Blair and Sharp, 1995). Our decoding approach suggests an alternative interpretation: anticipatory firing improves the representation of the present HD by compensating the motion-induced bias inherent in the population vector. If cells anticipate the HD by 25 ms, as observed in the ADN, the HD compass read-out improves by up to 40%. Moreover, across the layers within the HD circuit, the anticipatory time intervals (ATIs) should be staggered to optimize the population-vector codes, which is consistent with experimental observations (see, e.g., Taube, 2007). Similar phenomena are to be expected in other brain areas that encode stimuli that vary smoothly in time, neural coding will benefit from anticipation wherever sensory predictions or motor errors (see, e.g., Mumford, 1992; Wolpert et al., 1995; Berry et al., 1999; Rainer et al., 1999; Bastian, 2006; Bastos et al., 2012; Palmer et al., 2015) need to be calculated.

Materials and Methods

HD trajectories. The HD trajectories $\theta(t)$ were recorded in the Buzsáki laboratory (Mizuseki et al., 2009) and are publicly available at <http://dx.doi.org/10.6080/K0Z60KZ9>. The tracking data (file extension *.whl) are from male rats foraging on square platforms (side length: 120 or 180 cm).

The files contain time traces (sampling rate: 39.6 Hz) of the x and y coordinates of two LEDs mounted on the rats' head stage. To facilitate analyses at a fine temporal scale, we linearly interpolated the line segments between the data points to obtain an artificial sampling rate of 1 kHz. The time course of the angle between the vector connecting the LEDs (caudal to rostral) and the positive x axis defines the HD trajectory $\theta(t)$ with $-\pi \leq \theta(t) < \pi$. To avoid erroneous data points, we cut each tracking file at time points where tracking data were missing and only kept trajectory segments during which the distance between the LEDs was between 5 and 16 cm. Lower distances result from strong head tilts (e.g., during eating or rearing); larger distances indicate recording errors (distances cannot be larger than the distance between the LEDs). From the remaining segments, only those >10 s were kept for further analysis. Two of those segments were sorted out manually because of recording irregularities. In sum, this procedure resulted in 1163 segments. The angular head speeds have mean values of 87.7 ± 19.1 deg/s, median values of 60.4 ± 16.5 deg/s, and maximal values of 652 ± 173 deg/s; the forward speeds have mean values of 12.3 ± 3.0 cm/s, median values of 9.9 ± 2.9 cm/s, and maximal values of 65 ± 20 cm/s. The correlation between angular head speed and forward speed is 0.45, indicating that faster head turns are more likely during periods of faster forward motion.

A representative segment with a duration T_E of approximately 90 s (see Fig. 3) was chosen for the Monte Carlo simulations described below. This segment length is a compromise between sampling enough of the (foraging) behavior of the animal while keeping the computation time at a moderate level, and leads to representative statistics as shown in Figure 7. We mainly use the radian measure for angles in Materials and Methods. In the figure legends and Results, we state the angles in the degree measure for a more intuitive reading.

For comparison, we analyzed a second set of movement data from foraging male rats (Sargolini et al., 2006). As for the Mizuseki et al. (2009) data, we used recordings with two LEDs and selected trajectory periods by excluding erroneous data points, periods of very short and very long LED distances, and all short trajectory segments between these points. The Sargolini et al. (2006) data are already smoothed with a 300 ms running-average filter to remove tracking jitter. When the same smoothing operation is applied to the Mizuseki data, the difference of the duration-weighted averages between both datasets is $<0.15^\circ$ for time windows $T < 250$ ms (for mathematical definitions, see below). This indicates that the Mizuseki et al. (2009) dataset and the trajectory that we selected from this dataset for our simulations capture representative movement statistics of a rat during foraging behavior.

To analyze the effect of anticipatory firing, the time-dependent angular head velocity (AHV) $\omega(t)$ needs to be known. We numerically calculated $\omega(t)$ from the experimentally obtained HD trajectories $\theta(t)$ by central differences (i.e., $\omega(t_i) = [\theta(t_{i+1}) - \theta(t_{i-1})]/[t_{i+1} - t_{i-1}]$) to make sure that position and velocity samples are synchronized. In this case, the circular trajectory $\theta(t)$ was unfolded at the wrap-around $\theta(t) = \pm\pi$.

Spike train model. Each model neuron j generates an inhomogeneous Poisson spike train with time-dependent rate $\lambda_j(\theta(t))$. Accordingly, the number of spikes $k_{j,T}(t)$ fired in a time window of length T ending at time t is given by a Poisson distribution with time-dependent mean $\alpha_{j,T}(t) = \int_{t-T}^t \lambda_j(\theta(s)) ds$ as follows:

$$P(k_{j,T}(t)) = \frac{[\alpha_{j,T}(t)]^{k_{j,T}(t)}}{k_{j,T}(t)!} e^{-\alpha_{j,T}(t)} \quad (1)$$

The population response in that time window is given by the spike-count vector $\vec{K}_T(t) = (k_{1,T}(t), k_{2,T}(t), \dots, k_{N,T}(t))$. Assuming independence, the probability $P(\vec{K}_T(t))$ to observe a specific spike-count vector factorizes (i.e., $P(\vec{K}_T(t)) = \prod_{j=1}^N P(k_{j,T}(t))$). Figure 3 shows a raster plot of HD spike trains generated by this process.

Homogeneous and inhomogeneous cell populations. Our mathematical analysis focuses on homogeneous populations of HD cells. In these model populations, neurons have identical von Mises tuning curves $\lambda_j(\theta)$, given by Equation 17 in Results, which differ only by their equidistant preferred directions θ_j . All other parameters (maximum firing rate f_{max} , background firing rate f_{bg} , inverse tuning width κ , and ATI τ)

are chosen as the mean values from Table 1, which contains average experimental HD tuning parameters. The mean (background-corrected) population firing rate is:

$$R = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{i=1}^N [\lambda(\theta - \theta_i) - f_{bg}] d\theta = N[f_{max} - f_{bg}]e^{-\kappa} I_0(\kappa),$$

where $I_0(\kappa)$ is the modified Bessel function of first kind. Using $I_0(\kappa) \approx e^{\kappa}/\sqrt{2\pi\kappa}$, one obtains (Zhang, 1996) the following:

$$R \approx N[f_{max} - f_{bg}]/\sqrt{2\pi\kappa} \quad (2)$$

The analytic results obtained for homogeneous populations serve as a good approximation, also in the case of the related inhomogeneous populations (especially as $N \rightarrow \infty$; see Results).

For a description that is as close as possible to the biological findings, we simulate inhomogeneous populations. In such populations, every neuron has individual tuning parameters ($f_{max,j}$, $f_{bg,j}$, κ_j and τ_j), as shown by the example in Figure 3. Each of these parameters is sampled from a beta distribution, that is, a probability density function $p(x) \propto (x - x_{min})^{a-1}(x_{max} - x)^{b-1}$ on the interval $[x_{min}, x_{max}]$, whose parameters a and b are obtained from fitting the distribution's mean and variance to the values shown in Table 1 (see also Fig. 2). By shuffling the background firing rates, we ensure that the ratio between a cell's maximal firing rate $f_{max,j}$ and its background firing rate $f_{bg,j}$ exceeds 5 for every cell, as suggested by Taube (2007).

This method is not meant to exactly replicate the experimental tuning parameter distributions; rather, we use the method to explore the qualitative effect of inhomogeneous cell properties on the HD code. Nevertheless, because the first two moments of the experimental distributions as well as the observed parameter ranges are reproduced by the beta distributions, they mimic realistic HD populations. A comparison of Figure 2 with Taube (1995, their Figs. 5, 6) shows that the model distributions capture key features of the experimental histograms; however, our distributions have less pronounced peaks, especially for f_{max} and f_{bg} . To model the preferred HD of each cell, an equidistant arrangement ($\theta_j = j \times 2\pi/N - \pi$) is distorted by adding to each θ_j a random shift that is chosen from a Gaussian distribution with mean 0 and SD $2\pi/N$.

Circular error measure. The accuracy of the HD compass is determined by taking the angle $\hat{\theta}(t)$ of the population vector (see Eq. 17) and computing its average deviation from the true trajectory $\theta(t)$. Because we are calculating a difference of circular directions, which are defined on the interval $[-\pi, \pi)$, measures such as the squared difference between $\hat{\theta}(t)$ and $\theta(t)$ are not appropriate. Consider, for example, cases in which $\hat{\theta}(t)$ and $\theta(t)$ represent similar directions but lie at different sides of the wrap-around at $\pm 180^\circ$, for example, $\hat{\theta}(t) = 175^\circ$ and $\theta(t) = -175^\circ$. Although the two directions actually differ by only 10° , the squared difference measure would be $(350^\circ)^2$. A well-defined error measure for this situation is the circular distance given by $1 - \cos(\hat{\theta}(t) - \theta(t))$, as discussed by Jammalamadaka and Sengupta (2001). For small angular deviations, the circular distance approaches the squared difference, as illustrated in Figure 5A:

$$\begin{aligned} 1 - \cos(\hat{\theta}(t) - \theta(t)) &\approx \frac{1}{2}(\hat{\theta}(t) - \theta(t))_{loc}^2 \\ &\equiv \min_{k \in \{0, \pm 1\}} \left\{ \frac{1}{2}(\hat{\theta}(t) - \theta(t) - 2k\pi)^2 \right\} \end{aligned} \quad (3)$$

Averaging the circular distance between the population vector's direction and the true trajectory value over all possible neural responses defines the mean circular error as follows:

$$D(t) = E[1 - \cos(\hat{\theta}(t) - \theta(t))] \quad (4)$$

where the expectation is taken over all realizations of the spike-count vector. The mean circular error is the circular equivalent of the mean squared error used for nondirectional variables. Averaging the mean

circular error $D(t)$ along the underlying trajectory $\theta(t)$ results in the average mean circular error:

$$D = \frac{1}{T_e - T} \int_T^{T_e} D(t) dt \quad (5)$$

The error measure D includes the influences of both the stimulus dynamics and the noisy spike trains on the accuracy of the HD code. Compared with studies on the encoding and decoding of static stimuli, such as Mathis et al. (2012), the trajectory takes the role of the prior stimulus distribution. For graphical representation, we transform the error D to the degree-scale by defining the estimator's accuracy $A = (180^\circ/\pi) \cdot \arccos(1 - D)$, similar to the root mean square error used to describe the performance when dealing with noncircular stimuli.

Properties of the circular error measure. Because the reader might not be familiar with directional statistics, we briefly introduce the equivalents of the mean value, variance, and bias (see, e.g., Mardia and Jupp, 2009). The equivalent of the mean value is the mean or "expected" direction $\langle \hat{\theta}(t) \rangle = \arg\{E[e^{i\hat{\theta}(t)}]\}$. Note that the *arg* operation is performed *after* the expectation E is taken, again to deal with the circular wrap-around in a well-defined manner.

The circular variance is given as $V[\hat{\theta}(t)] = 1 - |E[e^{i\hat{\theta}(t)}]| = 1 - E[\cos(\hat{\theta}(t) - \langle \hat{\theta}(t) \rangle)]$. Accordingly, the mean circular error can be decomposed into contributions of the circular variance and a circular bias:

$$D(t) = V[\hat{\theta}(t)] + \{1 - V[\hat{\theta}(t)]\} \cdot [1 - \cos(\langle \hat{\theta}(t) \rangle - \theta(t))], \quad (6)$$

where the second term captures the systematic deviation of the estimator's mean direction from the actual trajectory value. This latter term is thus the circular equivalent of a bias (for a visualization of the two error sources, see Fig. 5) and Equation 6 is the circular equivalent of the decomposition of the mean square error into variance and squared bias.

Treatment of the zero spike-count case. If not a single HD cell fires in the interval $[t-T, t]$ (i.e., $\tilde{K}_T(t) = \vec{0}$), the direction of the population vector is undefined and no information about the current HD is available. The probability for this case can be approximated by $P_{\vec{0},T} = P(\tilde{K}_T(t) = \vec{0}) \approx e^{-(R+Nf_{bg})T}$, where the population firing rate R is defined in Equation 2. To extend the decoding scheme in a consistent manner to this case, $\hat{\theta}(\vec{0})$ should be picked randomly from $U_{[-\pi,\pi)}$, the uniform distribution on the interval $[-\pi, \pi)$. Separating this special case from all $\tilde{K} \neq \vec{0}$, one obtains the following:

$$\begin{aligned} D(t) &= P_{\vec{0},T} E[1 - \cos(U_{[-\pi,\pi)} - \theta(t))] \\ &+ \sum_{\tilde{K}_T \neq \vec{0}} (1 - \cos(\hat{\theta}(\tilde{K}_T) - \theta(t))) P(\tilde{K} = \tilde{K}_T(t)) \end{aligned} \quad (7)$$

The size of the term $P_{\vec{0},T} = e^{-(R+Nf_{bg})T}$ determines whether the case $\tilde{K}_T = \vec{0}$ is relevant or not. For the mean parameters of the ADN HD system shown in Table 1, $T \sim 1/(10\text{Hz} \cdot N)$ in order for the zero-spike case to contribute significantly. Even for a small population of only $N = 100$ neurons, this time scale is not more than 1 ms, much smaller than the optimal read-out times, in which we are interested. This shows that the zero-spike case is not relevant for our study so that we will not include the extended definition of the error in our analytical calculations.

Causal population-vector code. Equation 19 in the main text describes the relation $\langle \hat{\theta}_T(t) \rangle \approx \bar{\theta}_T(t)$ between the expected HD of the population vector and the corresponding time-averaged true HD (for the mathematical definition of these quantities, see Results). We now derive this relation for two different scenarios; both require homogeneous populations, but for large enough N , the results are also useful to describe inhomogeneous populations. In the context of 2D place fields and a maximum likelihood decoder, Mosheiff et al. (2017) recently derived a similar relationship.

Let us first assume that the encoded trajectory segments have constant AHV ω . This allows us to exploit a symmetry that arises for trajectory segments that are centered around the preferred direction of one partic-

ular neuron. For notational simplicity, let us assign the index 0 to this neuron and $\theta_0 = 0$, so that $\theta(t) = \omega \cdot \left(t - \frac{T}{2}\right)$ with $t \in [0, T]$. The other neurons are indexed such that $\theta_{-j} = -\theta_j$. This arrangement implies that the activity of the population is symmetric around $\theta = 0$ in the time interval $t \in [0, T]$, i.e., $\alpha_{-j}(t = T) = \alpha_j(t = T)$. Introducing the notation $\tilde{K}'$ for a count vector with mirror indices relative to $\tilde{K}$, that is, $k_{j'} = k_{-j}$, the symmetric activation implies $P(\tilde{K}'_T(t = T)) = P(\tilde{K}_T(t = T))$, while $\hat{\theta}_T(\tilde{K}') = -\hat{\theta}_T(\tilde{K})$. Using these identities, the mean direction $\langle \hat{\theta}_T(t = T) \rangle$ of the estimator $\hat{\theta}_T(t = T)$ is given by:

$$\begin{aligned} \langle \hat{\theta}_T(t = T) \rangle &= \arg \left\{ \frac{1}{2} \sum_{\tilde{K}} \left[e^{i\hat{\theta}(\tilde{K})} P\left(\tilde{K}\left(t = \frac{T}{2}\right)\right) + e^{i\hat{\theta}(\tilde{K}')} P\left(\tilde{K}'\left(t = \frac{T}{2}\right)\right) \right] \right\} \\ &= \arg \left\{ \frac{1}{2} \sum_{\tilde{K}} [e^{i\hat{\theta}(\tilde{K})} + e^{-i\hat{\theta}(\tilde{K})}] P\left(\tilde{K}\left(t = \frac{T}{2}\right)\right) \right\} \\ &= \arg \left\{ \sum_{\tilde{K}} \cos(\hat{\theta}(\tilde{K})) P\left(\tilde{K}\left(t = \frac{T}{2}\right)\right) \right\} = 0 = \bar{\theta}_T(t = T). \end{aligned} \quad (8)$$

An illustration of the symmetry of the distribution of $\hat{\theta}(t = T/2)$ around the trajectory average of a trajectory segment with constant angular velocity is shown in Figure 6A.

For an encoded variable that is static, the mean population vector can be calculated following Glasius et al. (1997). Applying their approach to a general, time-varying HD signal $\theta(t)$, we obtain the following:

$$\begin{aligned} E \left[\sum_{j=1}^N k_{j,T}(t) e^{i\theta_j} \right] &= \sum_{j=1}^N \left[\int_{t-T}^t \lambda_j(\theta(s)) ds \right] e^{i\theta_j} \\ &= \frac{N}{2\pi} \int_{t-T}^t \sum_{j=1}^N \lambda(\theta_j - \theta(s)) e^{i\theta_j} \frac{2\pi}{N} ds \\ &\approx \frac{N}{2\pi} \int_{t-T}^t \int_{-\pi}^{\pi} \lambda(\theta - \theta(s)) e^{i\theta} d\theta ds \\ &= \frac{N}{2\pi} \int_{t-T}^t \int_{-\pi}^{\pi} \lambda(\phi) e^{i(\phi + \theta(s))} d\phi ds \\ &= \frac{N}{2\pi} \int_{-\pi}^{\pi} \lambda(\phi) e^{i\phi} d\phi \int_{t-T}^t e^{i\theta(s)} ds \\ &= \frac{NC}{2\pi} \int_{t-T}^t e^{i\theta(s)} ds \\ \text{with } C &= \int_{-\pi}^{\pi} \lambda(\phi) e^{i\phi} d\phi \in \mathbb{R} \end{aligned} \quad (9)$$

so that

$$\arg \left\{ E \left[\sum_{j=1}^N k_{j,T}(t) e^{i\theta_j} \right] \right\} = \arg \int_{t-T}^t e^{i\theta(s)} ds = \bar{\theta}_T(t) \quad (10)$$

Thus, the mean population vector points into the direction of the trajectory average $\bar{\theta}_T(t)$. We cannot directly transfer this result to the mean direction $\langle \hat{\theta}_T(t) \rangle$ of the population vector, but it is reasonable to expect a similar behavior. Indeed, as shown by the example trajectory in Figure 6B, the numerical estimates agree well with $\langle \hat{\theta}(t) \rangle$. As also demonstrated

in this figure, the population vector effectively smoothens the representation of a trajectory and delays it by a time lag of $T/2$.

The histogram in Figure 6A shows that, for reasonable HD velocities and neural population sizes, the distribution of $\hat{\theta}(t)$ is narrow and closely concentrated around $\langle \hat{\theta}(t) \rangle$. We can thus expand the mean circular error $D(t)$ for small local angular distances to first nonvanishing order as $D(t) \approx \frac{1}{2} E \left[(\hat{\theta}(t) - \langle \hat{\theta}(t) \rangle)_{loc}^2 \right] + \frac{1}{2} (\langle \hat{\theta}(t) \rangle - \theta(t))_{loc}^2$. To obtain an approximation of the time-averaged circular error D , we integrate this expression along the whole trajectory over the corresponding time window i.e., $T_e - T$, where T_e indicates the trajectory's duration:

$$\begin{aligned} D &\approx \frac{1}{2(T_e - T)} \int_T^{T_e} \left\{ E \left[(\hat{\theta}(t) - \langle \hat{\theta}(t) \rangle)_{loc}^2 \right] + (\langle \hat{\theta}(t) \rangle - \theta(t))_{loc}^2 \right\} dt \\ &\equiv \frac{1}{2}(V + B^2). \end{aligned} \quad (11)$$

V and B^2 denote the (approximate) time-averaged variance and time-averaged squared bias, respectively.

Figure 6 also shows that the width of the $\hat{\theta}(t)$ distribution is only mildly affected by the AHV, even for high velocities. This implies that its variance V depends only weakly on time t so that we can omit the temporal average when calculating V . Furthermore, as the $\hat{\theta}(t)$ distribution hardly varies with ω , results by Seung and Sompolinsky (1993) for the (local) variance of the direction of the population vector (derived assuming a static stimulus) apply. We use their approach, which is based on calculating the Fisher information, to approximate the local squared deviation of the population vectors angle from its mean direction as follows:

$$E \left[(\hat{\theta}_T(t) - \langle \hat{\theta}_T(t) \rangle)_{loc}^2 \right] \approx E[(\hat{\theta}_T - \theta)^2] = \frac{1}{2NT} \frac{\bar{\lambda}_0 - \bar{\lambda}_2}{\bar{\lambda}_1^2} \quad (12)$$

where θ is a static stimulus parameter, $\hat{\theta}_T$ its estimator, and

$$\bar{\lambda}_n = \frac{1}{2\pi} \int_0^{2\pi} \lambda(\theta) e^{in\theta} d\theta$$

are the Fourier components of the zero-centered tuning curve $\lambda(\theta)$. This result holds for any symmetric unimodal tuning curve.

For von Mises curves (2), one obtains $\bar{\lambda}_n = (f_{max} - f_{bg})e^{-\kappa} I_n(\kappa) + f_{bg}\delta_{0n}$ because $I_n(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(\theta)} \cos(n\theta) d\theta$. Inserting the first two terms of the expansion of each Bessel function for large κ , we arrive at the following:

$$V \approx E \left[(\hat{\theta}_T(t) - \langle \hat{\theta}_T(t) \rangle)_{loc}^2 \right] \approx \frac{1 + \frac{f_{bg}}{f_{max} - f_{bg}} \sqrt{\kappa^3 \pi/2}}{\kappa RT \left(1 - \frac{3}{4\kappa}\right)} \quad (13)$$

for a homogeneous neural ensemble with von Mises tuning curves (see Fig. 7, dotted lines). This shows that, as the average number RT of (non-background) spikes in the spike-count window increases, the variability V decreases and more information about the stimulus is available. Similarly, the variance of the estimator increases for higher background activity (Seung and Sompolinsky, 1993; Mathis et al., 2012).

To first approximation, the time-averaged circular error can be written as

$$D \approx \frac{1 + \frac{f_{bg}}{f_{max} - f_{bg}} \sqrt{\kappa^3 \pi/2}}{2\kappa RT \left(1 - \frac{3}{4\kappa}\right)} + \frac{1}{2(T_e - T)} \int_T^{T_e} (\bar{\theta}_T(t) - \theta(t))_{loc}^2 dt \quad (14)$$

Figure 1. Population encoding and decoding of HD. **A**, The HD system acts as an internal compass, encoding the time-dependent azimuth $\theta(t)$ of the animal's head with respect to a landmark-anchored reference direction. **B**, Polar plot of idealized HD tuning curves and activities. HD cells are broadly tuned ($\sim 90^\circ$) so that $\sim 25\%$ of all cells are active at any moment. The tuning curves, highlighted for 20 neurons by the inner blue and red lines, have an elevated baseline (innermost circle), and their maximum is achieved at the cell's preferred direction, as illustrated for the red cell. The activation levels of the different cells at a particular time t are denoted by the disk radii. The population vector $\hat{\theta}(t)$ is calculated from these stochastic activities and provides the time-dependent direction estimate. **C**, Temporal evolution of the HD activity packet for a small clockwise head movement $\theta(t)$, shown as a dashed line. Blue, red, and gray dots in the raster plot represent spikes generated with inhomogeneous Poisson processes whose rates are determined by the current HD via the tuning curves shown on the left. Model parameters: peak firing rate, $f_{max} = 50$ Hz; background firing rate, $f_{bg} = 2$ Hz.

Figure 2. Tuning curves: shapes and parameter ranges. **A**, Left, Middle, Illustration of von Mises tuning curves for different choices of the minimum, mean, and maximum values of the peak firing rate f_{max} and tuning width σ (for anticipatory time delay $\tau = 0$). Right, Shift of a tuning curve during fast head turns (AHV: $\pm 360^\circ/s$); note the different scale on the θ -axis. **B**, Parameter distributions from which the tuning curves were sampled to generate inhomogeneous populations (beta distributions whose ranges were adapted to the experimental observations). The distributions have means, SDs, and range values are as shown in Table 1.

where the second term needs to be evaluated for the encoded trajectory $\theta(t)$. In Figure 7 (right column), Equation 14 is shown as a function of T using the representative trajectory segment $\theta(t)$ mentioned above. The tuning curve parameters are chosen according to the mean values shown in Table 1.

Monte Carlo simulations. We use Monte Carlo simulations of the average circular error to check the analytical results and assess the effect of inhomogeneous tuning parameters. To this end, we sample the integral over the trajectory and the Poisson noise for $M \gg 1$ times: first, an endpoint of a trajectory segment $t_i \in [T, T_e]$ is randomly selected. Then, a count vector $\tilde{K}_i = \tilde{K}(t_i)$ is sampled from the respective Poisson distribution. Finally, the circular distance $d(t_i) = 1 - \cos(\hat{\theta}(\tilde{K}_i) - \theta(t_i))$ is calculated and the average circular error is approximated as $\hat{D} = 1/M \sum_{i=1}^M d(t_i)$.

Experimental design and statistical analysis. We reanalyzed data originally recorded by Mizuseki et al. (2009) and by Sargolini et al. (2006), and we refer the reader to those publications for details on the experimental designs. All our analyses were performed in Python (RRID:SCR_008394) using functions from the SciPy (RRID:SCR_008058) and NumPy (RRID:SCR_008633) modules.

Results

The HD system of rodents encodes the time-dependent azimuth $\theta(t)$ of the animal's head relative to a landmark-anchored refer-

ence direction (Fig. 1) and plays a functional role for spatial navigation (Butler et al., 2017). Notably, during head turns, HD cells start to fire before their preferred direction θ_j is reached, as though the neurons anticipated the future HD (Blair and Sharp, 1995). Faster rotations lead to stronger shifts in time than slow turns. Denoting the time-dependent AHV by $\omega(t)$, the shift can be approximated by the following:

$$\theta_j \rightarrow \theta_j - \omega(t) \cdot \tau_j \quad (15)$$

where the time constant τ_j is called the ATI of cell j and describes the extent to which HD firing seems to be shifted forward in time (Blair and Sharp, 1995; Blair et al., 1997; Taube and Muller, 1998); for a visualization, see Fig. 2A, right.

The rodent HD system consists of multiple processing stages along which ATIs decrease from stage to stage: from the lateral mammillary nuclei (LMN, $\tau \approx 40$ – 75 ms) to the anterodorsal thalamic nucleus (ADN) and retrosplenial cortex ($\tau \approx 25$ ms) to the postsubiculum, where neurons tend to neither lag nor anticipate the cell's preferred firing direction (Taube, 2007). As with other HD cell parameters (see also Table 1), ATIs vary from cell to cell (Blair et al., 1997; Taube and Muller, 1998). At the mechanis-

Figure 3. Encoding model and HD trajectory. **A**, Tuning curves, modeled as von Mises functions, of an inhomogeneous population of HD cells in the ADN ($N = 100$). The parameters of the von Mises functions were drawn from the distributions shown in Figure 2. **B**, Time-dependent HD $\theta(t)$ obtained from a two LED recording of a foraging rat (black line). In addition, spike times from one realization of inhomogeneous Poisson processes based on the rat's movement trajectory are shown for each neuron.

tic level, ATIs have been explained by an interplay between afferent dynamics and movement statistics (see also Zhang, 1996; van der Meer et al., 2007; Tsanov et al., 2014), but a consistent functional interpretation of ATIs is still missing.

As we will show, anticipatory firing improves the accuracy with which the present HD can be decoded from time-dependent HD signals. For concreteness, we use the population-vector decoder (Georgopoulos et al., 1983, 1986), which provides a reliable estimate of variables encoded by the distributed activity pattern of a neural ensemble (Seung and Sompolinsky, 1993; Salinas and Abbott, 1994; Glasius et al., 1997; Stemmler et al., 2015). We focus on HD cells in the ADN, which contains the highest fraction of HD cells (Taube and Bassett, 2003) and is considered a “core region” (Taube, 2007) or even “central hub” (Peyrache et al., 2017) of the HD system. To cover the dynamic character of head movements, we explicitly consider time-varying stimuli (Mosheiff et al., 2017). We take a computational approach and use experimentally recorded HD trajectories $\theta(t)$ to generate artificial spike trains (Fig. 3). The periodic tuning curves of individual HD cells can be captured by von Mises functions (Zhang, 1996) whose tuning width, peak firing rate, and background firing rate are chosen according to the available experimental literature (Fig. 2B). Spike generation is modeled by inhomogeneous Poisson processes (see Eq. 1). From the spike trains of this simulated neural population, we compute an estimated trajectory $\hat{\theta}(t)$ based on the population-vector decoder. We then measure the accuracy of the HD system by analyzing the distribution of $\hat{\theta}(t)$ and comparing its mean $\langle \hat{\theta}(t) \rangle$ with the true trajectory $\theta(t)$, as described in Materials and Methods. Using this framework, we systematically explore the HD representation and characterize the effect of anticipatory firing. We derive an analytic approximation for the population code's accuracy and use numerical simulations to check its validity and to study populations with inhomogeneous tuning curves. An approach based on a reconstruction from experimentally measured spike trains, as in Johnson et al. (2005) for the postsubiculum, is currently not feasible because simultaneous recordings of HD cells in ADN are only available for small populations of less than 20 HD neurons (Peyrache et al., 2015).

HD trajectories

The HD trajectories $\theta(t)$ for this study are taken from a dataset that was recorded in the laboratory of G. Buzsáki (Mizuseki et al., 2009). Because the HD is a circular variable, temporal averages

and expectation values need to be consistent with circular statistics (see Materials and Methods). In particular, the temporal average $\bar{\theta}_T(t)$ of a trajectory segment $\theta(s)$ within some time interval $[t - T, t]$, is defined by:

$$\bar{\theta}_T(t) = \arg \left\{ \frac{1}{T} \int_{t-T}^t e^{i\theta(s)} ds \right\} \quad (16)$$

so that for constant AHV (i.e., $\theta(t) = \theta_0 + \omega \cdot t$) the trajectory average corresponds to the midpoint of the trajectory segment, $\bar{\theta}_T(T/2) = \theta_0$. To avoid overly complicated mathematical expressions, we will omit indices, such as T in $\bar{\theta}_T(t)$, or variables, such as t in $\bar{\theta}_T(t)$, unless this could cause misunderstanding.

Tuning curves and spike-train model

HD cells in the ADN have been studied in great detail (see e.g., Knierim et al., 1995, 1998; Blair and Sharp, 1995; Blair et al., 1997, 1998; Goodridge and Taube, 1997; Stackman and Taube, 1998; Taube and Muller, 1998; Taube, 2007; Clark et al., 2009, 2012; Peyrache et al., 2015, 2017; Tan et al., 2015). Their stereotypical tuning profiles show a single hump and are well fit by von Mises functions (Zhang et al., 1998):

$$\lambda_j(\theta) = (f_{\max,j} - f_{bg,j}) \cdot e^{\kappa_j[\cos(\theta - \theta_j) - 1]} + f_{bg,j} \quad (17)$$

with peak firing rate $f_{\max,j}$, background firing rate $f_{bg,j}$, “concentration” parameter κ_j , and preferred direction θ_j . The cell's tuning width σ_j can be calculated as $\sigma_j = 180^\circ / (\pi \sqrt{\kappa_j})$. Based on these tuning curves, each model neuron generates an inhomogeneous Poisson spike train (see Eq. 1) with time-dependent rate $\lambda_j(\theta(t))$. If we use λ without subscript, we refer to a neuron with preferred direction $\theta = 0$.

According to Taube and Bassett (2003), there are $N \approx 12,000$ HD cells in ADN per brain hemisphere. We therefore study ensembles up to this size but also consider smaller N to analyze size-dependent effects.

The tuning parameters of HD cells vary strongly from neuron to neuron. To evaluate the effect of such variations on the neural code, we describe the parameters by probability distributions whose ranges, means, and SDs are estimated from the literature (see Table 1; Materials and Methods). Sampling the parameters of each model neuron from these distributions results in what we call an inhomogeneous population (Fig. 3). To model the preferred HD of each cell, the equidistant arrangement

Figure 4. Causal decoding and effect of the spike-count window. **A**, Within a causal framework, only spikes from the past can be used to estimate the encoded variable. To decode the time-dependent HD $\theta(t)$, shown as dashed black line at the time of the solid black line, one might, for example, rely on spikes generated within the gray area. The blue line indicates the time-dependent estimate $\hat{\theta}(t)$, which was calculated from the shown spike responses using the population-vector decoder. **B**, **C**, Estimated HDs $\hat{\theta}(t)$ using a short (**B**) and a long (**C**) spike-count window T . For the short spike-count window, $\hat{\theta}(t)$ jitters strongly around the true value $\theta(t)$, whereas for longer T the estimated HD is much smoother but lags behind the true trajectory because spikes from the more distant past are used.

$$\theta_j = j \times 2\pi/N - \pi$$

is altered by adding to each θ_j a random shift chosen from a Gaussian distribution with mean 0 and SD $2\pi/N$. The mathematical analysis focuses on homogeneous populations. Here, neurons differ only by their (equidistant) preferred direction but share all other parameters (so that we omit the subscript j), chosen as the mean values in Table 1. For simulations of inhomogeneous populations, we sample ATIs from the distribution shown in Figure 2; for homogeneous populations, we assume $\tau = 25$ ms for all cells, unless noted otherwise.

Causal population vector reconstruction

The population vector $PV(t)$ is a weighted sum, $PV(t) = \sum_{j=1}^N k_{j,T}(t) V_j$, where the V_j are unit vectors in θ_j direction and $k_{j,T}(t)$ denotes the number of spikes fired between time $t - T$ and t (Georgopoulos et al., 1983, 1986). Introducing the phasors $\exp(i\theta_j) = \cos(\theta_j) + i \sin(\theta_j)$, the population vector can be written as $\sum_{j=1}^N k_{j,T}(t) \cdot e^{i\theta_j}$; its direction is given by the argument of this complex number. Defining the spike-count vector $\tilde{K}_T(t) = (k_{1,T}(t), k_{2,T}(t), \dots, k_{N,T}(t))$, the animal's HD can be decoded from the experimental or modeled HD cell responses as follows:

$$\hat{\theta}_T(t) \equiv \hat{\theta}(\tilde{K}_T(t)) = \arg\left\{\sum_{j=1}^N k_{j,T}(t) \cdot e^{i\theta_j}\right\} \quad (18)$$

The decoder $\hat{\theta}_T(t)$ is chosen to obey causality: only spikes from spike-count windows ending at time t are used to estimate $\theta(t)$, as illustrated in Figure 4. This choice reflects the fact that any downstream processing step of the HD signal can only depend on spikes that were fired in the past.

Read-out accuracy: a function of systematic lag (bias) and random noise (variance)

The theoretical framework now at hand allows us to calculate how accurately the current HD $\theta(t)$ can be determined from the time-dependent population activity of non-anticipatory HD cells in a moving animal. Because the PV estimator $\hat{\theta}_T(t)$ relies exclusively on spikes fired before time t , the computed HD always lags behind the current HD $\theta(t)$. In statistical terms, this means that $\hat{\theta}_T(t)$ is biased with respect to $\theta(t)$. Additionally, because the spiking behavior of an ensemble of neurons is noisy, $\hat{\theta}_T(t)$ exhibits random fluctuations, which are measured by the variance V . As

shown in Materials and Methods, the (circular) error D for the HD read-out from a neural ensemble is given by the sum of the (squared) bias B and the variance V as follows:

$$D = \frac{1}{2}(V + B^2) \quad (19)$$

In the following, we describe the behavior of the bias B and the variance V of the causal PV estimator and demonstrate at which point anticipatory firing introduces an improvement of the read-out accuracy.

To determine the bias B of the PV estimator, we start by evaluating the estimator's average at a given instance of time. As derived in Materials and Methods,

$$\langle \hat{\theta}_T(t) \rangle \approx \bar{\theta}_T(t). \quad (20)$$

This means that, when averaging out the noise induced by stochastic spiking, the PV estimator provides the average HD of an animal during the spike-count time T . Because the bias is given by the difference between $\langle \hat{\theta}_T(t) \rangle$ and $\theta(t)$ (see Fig. 5), the dynamic nature of the HD induces a systematic error. To visualize this, consider a trajectory segment with constant AHV: $\theta(t) = \theta_0 + \omega t$. Here, $\langle \hat{\theta}_T(t) \rangle \approx \theta(t) - \omega T/2$, such that the bias amounts to $B = \omega T/2$. The PV estimator is thus biased for every spike-count window for which the average HD does not match the final HD. Generally, longer spike-count windows will lead to larger biases because the HD has more time to change so that the average HD deviates stronger from the endpoint HD (see also Figure 6). We evaluated the average squared bias B^2 along all experimentally recorded trajectories as a function of T . The results are shown in Figure 7A, and all share an increase with integration time T .

Based on this analysis of the bias, it appears reasonable to use very short spike count times to keep the spikes used for decoding as close as possible to the current time point t . This will, however, increase the noise of the read-out as less and less spikes are involved in the estimation and their stochastic variations cannot be averaged out anymore. Together, the two opposing effects define an optimal read-out time for which there is an ideal trade-off between systematic lag and statistical fluctuations.

The amount of error induced by noisy spiking (i.e., the variance V) can be approximated analytically using Fisher information. Because this noise component is independent of the AHV, we can refer to results derived for the well-studied static case (for

Figure 5. Circular error measure, bias, and variance. **A**, Blue line indicates circular distance function. For small circular distances, this function resembles the squared error (red dashed line). **B**, A total of 300 realizations of population vectors (blue arrowheads) calculated from a population of $N = 100$ neurons encoding a circular static variable θ . The mean direction $\langle \hat{\theta} \rangle$ represents the directional average of all possible realizations. If $\langle \hat{\theta} \rangle$ does not agree with the value of θ , which θ is supposed to represent, as illustrated in the figure, the estimator is called biased. The (circular) variance describes the directional extent of all possible realizations of the population vector given a certain encoded variable (the spread is caused by the noisy encoding).

the full result, see Materials and Methods; Eq. 13). We find that the variance decreases with the average number of collected spikes as $1/RT$, thus decreasing as spikes are collected during longer time windows (Fig. 7A, dotted lines).

According to Equation 12, the variance term depends only on the scaling factor $(2NT)^{-1}$ and the ratio $(\bar{\lambda}_0 - \bar{\lambda}_2)/\bar{\lambda}_1^2$. Comparing different tuning curves, normalized such that they result in the same population firing rate, numerical values for this ratio are 0.083 (von Mises tuning), 0.080 (Gauss tuning), and 0.085 (triangular tuning). This calculation shows that the variance term for von Mises tuning differs by less than 5% from the other two tuning functions often used for HD cells.

The trade-off arising from the two different error sources is in Figure 7 (right column). Along with data points obtained from applying our analytical treatment to an experimentally recorded trajectory, we show results from Monte Carlo simulations (see Materials and Methods). These involved either homogeneous or inhomogeneous HD cell ensembles and confirm our mathematical prediction. In particular, as N increases, results from inhomogeneous and homogeneous cell ensembles align because all possible HDs are uniformly covered by tuning curves.

When decoding time-independent stimuli, the read-out window T is usually assumed to be a free parameter that needs to be chosen in a biologically or methodologically useful way (see, e.g., Seung and Sompolinsky, 1993; Brown et al., 1998; Zhang and Sejnowski, 1999; Mathis et al., 2012). Considering the dynamic nature of a stimulus, as we do in this study, leads to specific spike-count windows T for which the highest possible accuracy is achieved (see also Mosheiff et al., 2017). The bias depends exclusively on the shape of the HD trajectory, whereas the variance depends exclusively on the parameters of the neural ensemble. For a non-anticipatory population of HD cells, the only chance to decrease the optimal read-out error is to invest more neural resources (N, f_{max}) and thus reduce the contribution of the random spiking noise to the error. A causal decoder cannot directly mitigate the systematic lag. We show in the following paragraph that anticipatory firing does influence the bias term and can lead to significant improvements of the HD code.

Effect of anticipatory firing in HD ensembles

As proposed by Blair and Sharp (1995), we assume that the shift of the preferred HD scales linearly with AHV. Writing $\tilde{\lambda}_j$ for the average firing rate of an anticipatory neuron in a homogeneous

population and inserting Equation 15 for the preferred direction, we obtain the following:

$$\begin{aligned}\tilde{\lambda}_j(\theta(t)) &= \lambda(\theta(t) - (\theta_j - \omega(t) \cdot \tau)) \\ &= \lambda((\theta(t) + \omega(t) \cdot \tau) - \theta_j) \\ &= \lambda_j(\theta(t) + \omega(t) \cdot \tau).\end{aligned}\quad (21)$$

This formulation treats anticipation in HD cells as a shift in the angular trajectory given by $\varphi(t) = \theta(t) + \omega(t)\tau$, as illustrated in Figure 8. If the AHV $\omega(t)$ were constant between t and $t + \tau$, the product $\omega(t)\tau$ equals the total turning angle in that time interval.

Accordingly, the expected direction of a homogeneous population of anticipatory neurons is given by the following:

$$\langle \hat{\theta}_{T,\tau}(t) \rangle \approx \bar{\varphi}_T(t) \quad (22)$$

The subscript τ indicates that we are describing an anticipatory HD population. Because the trajectory $\varphi(t)$ corresponds to an estimate of the future HD, this change of the expected direction $\hat{\theta}_{T,\tau}(t)$ can lead to a compensation of the bias of $\hat{\theta}(t)$ compared with the non-anticipatory case (Eq. 20).

To better understand this effect, we first examine trajectory segments with constant AHV ω (i.e., with $\theta(t) = \theta_0 + \omega \cdot t$). In this case, $\varphi(t) = \theta_0 + \omega \cdot (t + \tau) = \theta(t + \tau)$ so that $\varphi(t)$ is a time-shifted version of $\theta(t)$. We obtain

$$\langle \hat{\theta}_{T,\tau}(t) \rangle \approx \bar{\varphi}_T(t) = \bar{\theta}_T(t + \tau) = \theta(t + \tau)$$

for the expected direction of the causal estimator $\hat{\theta}(t)$. The ATI τ thus temporally shifts the expected direction of $\hat{\theta}(t)$ along the trajectory $\theta(t)$. Evaluation of the average bias term for a linear trajectory segment leads to $B = |\omega \cdot (T/2 - \tau)|$. Choosing τ within $(0, T)$ reduces the bias compared with the situation without anticipation. For constant AHV, it is even possible to fully remove the bias by choosing $\tau = T/2$, as illustrated in Figure 8.

Anticipation also reduces the bias when the velocity of the HD $\theta(t)$ varies in time. To quantify this effect in the experimentally relevant case, we recalculate B as for Figure 7A, but with anticipatory firing. We use Equation 21 to compute the expected HD. The results of this analysis are shown in Figure 7B, C (left panels) for $\tau = 25$ ms and $\tau = 50$ ms, respectively. In stark contrast to populations without anticipation (Fig. 7A), B does not increase monotonically with T but has a minimum. Similar to the case of trajectories with constant AHV, the average bias B is smallest for $T \approx 2\tau$. In this case, the implicit interpolation into the future described by $\varphi(t)$ has the optimal size. However, because of the time-varying nature of the AHV, the bias is not completely canceled.

In contrast, the variance of $\hat{\theta}(t)$ does not depend on the ATI τ , as seen by considering trajectory segments with constant AHV or the data shown in Figure 8. To describe the reconstruction error D for anticipatory firing analytically, we directly adopt the variance term from Equation 13 and replace the bias term by the anticipatory bias as analyzed above. We now numerically calculate the average circular error D with homogeneous and inhomogeneous populations. For the inhomogeneous populations, the ATIs are sampled from the distribution depicted in Figure 2 (for $\tau = 50$ ms, the distribution of τ is shifted by 25 ms to the right). Figure 7B, C (right panels) shows the analytical result from Equation 13 together with simulation results for ATIs of $\tau = 25$ ms (to mimic experimental ATIs) and $\tau = 50$ ms (to illustrate the effect of longer anticipation).

As in Figure 7A, the data shown in Figure 7B, C demonstrate a qualitative agreement between the analytic description and the

Figure 6. Relation between the population vector and the encoded trajectory. **A**, Distribution of the population vector's direction $\hat{\theta}(t)$ for short trajectory segments with constant angular velocities. The segments are shown below the histogram; their values at $t = 25$ ms are highlighted by large dots. The segment plotted with the red dotted line has $\omega = 360^\circ \text{ s}^{-1}$ and represents fast angular head velocities, as observed for foraging rats. The threefold faster yellow trajectory was chosen to generate a noticeable deviation of the histogram from the static (gray) case. To sample $\hat{\theta}(t = 25 \text{ ms})$, a homogeneous population with $N = 100$ neurons and time-dependent Poisson processes was used to generate spike-count vectors $\vec{K}$ (spike-count window $T = 50$ ms). Even for fast head turns, the distribution of $\hat{\theta}_T(t = 25 \text{ ms})$ is very similar to the distribution for the static case. In particular, the preferred direction for all three trajectory segments is at 90° , the trajectory average $\bar{\theta}(t = 25 \text{ ms})$, and the variance of the distribution increases only slightly with increasing ω . Regarding the trajectory endpoints, the estimator $\hat{\theta}_T(t = 25 \text{ ms})$ is thus biased for $\omega \neq 0$. **B**, In essence, the population vector provides a smoothed and delayed copy of the trajectory. Dashed black line indicates an artificial trajectory $\theta(t)$. Continuous lines indicate trajectory averages $\bar{\theta}_T(t)$ for various window lengths T . The mean direction of the population vector agrees well with the time-lagged trajectory averages (i.e., $\langle \hat{\theta}_T(t) \rangle \approx \bar{\theta}_T(t)$), even for trajectory segments with time-varying $\omega(t)$. The mean directions $\langle \hat{\theta}_T(t) \rangle$ are shown as dots. Error bars indicate the SD of the distribution from which they were computed.

simulations. Most notably, the results for populations with realistic ATI distribution align closely with the results for homogeneous populations (for sufficiently large N) and show that cell-to-cell ATI variability does not reduce the advantages of anticipatory firing. Compared with populations without anticipation, the minimal errors of anticipatory populations are shifted toward longer spike-count windows T because the bias does not increase monotonically with T , but rather has a minimum for $T > 0$. Consequently, spikes can be collected over longer time windows (as long as the bias remains small), which reduces the variance. For $\tau = 25$ ms, this improved trade-off leads to a reduction of the minimal error for every population size shown in Figure 7B. For large ATIs, however, the minimum achievable bias can become so high that the minimal reconstruction errors are larger than without anticipation (Fig. 7C). Hence, anticipation only improves the accuracy of the causal population-vector decoder if the ATIs lie within a certain interval.

In Figure 9A, we show how much the minimum achievable error can be reduced by anticipatory firing with different ATIs. To generate this figure, the function $A(T)$ was numerically minimized for $\tau = 0$ ms and various values $\tau > 0$ ms. The minimum errors for $\tau = 0$ ms are shown in the inset as a function of population size N . The main plot presents the (average) fraction of the minimum A with and without anticipatory firing. The figure reveals that anticipatory firing can lead to large improvements in reading out the HD compass: the experimentally determined average value of $\tau = 25$ ms increases the accuracy by up to 40%. Figure 9B demonstrates how large the population size N_0 of a non-anticipatory population has to be to reach the same accuracy as an anticipatory population of size N . For a large range of population sizes and ATIs, non-anticipatory populations need to be more than three times as large as populations of anticipatory neurons (Fig. 9B). For $\tau = 25$ ms, the ratio N_0/N is particularly high for (biologically plausible) population sizes between 1000

and 10,000 neurons. These results reveal that the experimentally reported value of $\tau = 25$ ms lies exactly in the range that leads to an optimal improvement of the HD code given the broadly tuned neurons and the temporal statistics of the head movements.

Discussion

Single HD cells in the rat ADN predict the animal's HD roughly 25 ms into the future (Blair and Sharp, 1995; Taube and Muller, 1998). Ironically, such anticipatory firing helps to encode the present HD, not the future HD, as we have shown here by analyzing the read-out process. Taking the temporal statistics of head movements in foraging rats into account, we found that the error in modeled neuronal population vectors is smallest when HD cells anticipate the true HD by an ATI of around 25 ms for populations with a few thousand neurons, in agreement with the ADN measurements. With realistic parameters for the HD tuning curves, anticipatory firing decreases the minimal error in the HD estimates by up to 40%. One would need a fivefold increase in the number of neurons to achieve the same performance in the absence of anticipation.

Because of their simplicity, population vectors are often the method of choice to decode the HD signal from multielectrode recordings (Peyrache et al., 2015) and calcium-imaging data (Kim et al., 2017). Moreover, population vectors are readily interpreted in the framework of ring-networks, which posits that vestibular and other sensory signals drive bumps of neuronal activity to move along a ring-like network of neurons (for review, see Laurens and Angelaki, 2018). Other decoding approaches, such as spike-by-spike population decoding (Huys et al., 2007), Bayesian filtering (Bobrowski et al., 2009), or predictive Kalman filters, are not likely to provide significant advantage over population-vector decoding. Indeed, in a similar task (i.e., the reconstruction of spatial trajectories from place-cell activity), such alternative strategies did not improve the decoding accuracy (Brown et al., 1998).

Various questions lie outside the scope of this study. We did not model path integration based on accumulating and error-correcting the HD signal (Cheung et al., 2008), nor did we address the mechanistic origin of anticipatory firing. Possible explanations range from neurons sensing head acceleration (Zhang, 1996) to high-pass filtering based on short-term synaptic depression (Puccini et al., 2007) or a combination of spike-rate adaptation and postinhibitory rebound (van der Meer et al., 2007). Instead, we focused on the role anticipatory firing plays in reducing the bias in estimating the time-varying HD signal.

Anticipation resolves a quandary inherent in decoding any time-dependent quantity from a population of spiking neurons. Given the stochastic nature of neuronal discharges, averaging over time should improve the decoding accuracy, were it not for the fact that the stimulus is continually changing. If the decoder reads out the population activity over some time window T , a trade-off arises (see, e.g., Brown et al., 1998; Zhang et al., 1998; Mosheiff et al., 2017; and our approach): the stimulus dynamics

Figure 7. Bias and accuracy of the population vector. **A**, Left, Mean bias B for all trajectory segments of the dataset (Mizuseki et al., 2009) as a function of segment duration T . The duration-weighted average of all trajectory segments is also shown. Error bars indicate the duration-weighted SD. Light and darker yellow areas represent the sets of trajectory segments between the 25th–75th and 5th–95th percentiles, respectively. Red dashed line indicates the representative trajectory, which was used as a basis of the simulations. For causal decoding without anticipation, the plotted data represent the best achievable accuracy for given T . Because longer time windows admit stronger changes of the HD, the bias term increases with increasing T . Right, Approximation results for the reconstruction accuracy given various population sizes (analytical results: continuous lines). For each population size, there is a spike-count window, which leads to an optimal accuracy. Larger populations can reach a higher HD accuracy. While the contribution of the variance to the total error decreases with the size of the population (dashed lines), the bias term (red dashed line) is independent of the population size. Dots represent simulation results from Monte Carlo simulations using homogeneous populations. Transparent lines indicate simulation results from different inhomogeneous populations (one line per population). As evident from the figure, the approximate solution derived in the text captures the population vector accuracy of homogeneous populations as well as of inhomogeneous populations (for large enough population size). **B**, **C**, Same as in **A**, but for ATIs $\tau = 25$ ms and $\tau = 50$ ms, respectively. Left panels, In both cases, the dip with smallest average bias B is found at $T \approx 2\tau$ because for these values the interpolated trajectory $\varphi(t)$ is closest to $\theta(t)$, on average. The length of the spike-count window with minimum bias is thus controlled by τ . Right panels, In addition to data for the nonzero τ , the analytic results from **A** are plotted as gray background lines. The simulation results (dots and transparent lines) again agree well with the analytic description (continuous, nontransparent lines), especially for large population sizes N . As the bias term has a minimum at $T > 0$, longer spike-count windows can be used compared with the nonanticipatory case. This reduces the variance and bias of the estimator and results in smaller minimal reconstruction errors. However, too long ATIs can also deteriorate population codes. One such example is shown for which the optimal accuracy of a population with $N = 10,000$ neurons is worse for anticipatory firing with $\tau = 50$ ms than for $\tau = 25$ ms.

causes the deviations in the population-vector read-out (relative to the true, instantaneous stimulus) to increase with T , whereas the variance for estimating a static stimulus would decrease with T (Burak and Fiete, 2012). Systematic biases caused by averaging over time are exacerbated by the requirement that a biologically plausible decoder should be causal: only past neural activity can be taken into account, not future activity. Temporal averaging will therefore shift the population vector to reflect an estimate of the stimulus that lies in the past, unless the neurons anticipated future stimulus values. In an ideal encoding of a time-varying stimulus (HD is just one example), the average ATI should be less

than the correlation time of the stimulus, whereas T , which reflects the effective integration time constant of a read-out, should be set to achieve the desired accuracy, given the number of neurons and their tuning properties. Crucially, though, T and τ should be matched, so that $T = 2\tau$, to make the population-vector estimate unbiased. This theoretical result applies to any neural system encoding stimuli that are gradually changing over time.

As shown in this study, measured ATIs in the rat ADN are ideally suited to maximize the read-out accuracy, given the animal's movement statistics, single-neuron parameters, and network size. ATIs are advantageous because the HD

Figure 8. Effects of anticipatory firing on population-vector decoding. Effect of anticipatory firing on the distribution of $\hat{\theta}(t)$ and its bias. **A**, Bottom row represents the three trajectory segments with constant AHVs ω (0, $360^\circ/\text{s}$, $3 \times 360^\circ/\text{s}$) used for the simulations (as in Fig. 6). Three top rows represent distribution of $\hat{\theta}(t)$ for $\tau = 10 \text{ ms}$, $\tau = 25 \text{ ms}$, and $\tau = 50 \text{ ms}$, respectively. An ATI of $\tau = 10 \text{ ms}$ leads to a reduction of the bias compared with the nonanticipatory case shown in Figure 6. The bias is reduced as the expected directions $\langle \hat{\theta}(t) \rangle = 25 \text{ ms}$ (vertical lines) are closer to the trajectory's endpoints (depicted as dots). For $\tau = 25 \text{ ms}$, the condition $\tau = T/2$ (see main text) is fulfilled, and the expected directions agree exactly with the trajectory's endpoints. In this situation, $\hat{\theta}_T(t)$ is an unbiased estimator of $\theta(t)$. For $\tau = 50 \text{ ms}$, the bias is as large as in the nonanticipatory case, but with opposite orientation. The width of the distributions does not depend on the ATI. **B**, Raster plots of 100 HD cells encoding the trajectory segment $\theta(t)$ without anticipation ($\tau = 0$). Blue line indicates the decoded trajectory $\hat{\theta}(t)$, which lags behind the true trajectory $\theta(t)$. **C**, With an anticipation of $\tau = 25 \text{ ms}$, the decoded trajectory no longer lags behind the true trajectory. Red line indicates the modified trajectory $\varphi(t) = \theta(t) + \omega(t)\tau$. **D**, Same as in Figure 6, but for anticipatory firing with $\tau = 25 \text{ ms}$. Continuous lines indicate the lagging circular averages of the modified trajectory $\varphi(t)$. They agree well with the mean directions $\langle \hat{\theta}_T(t) \rangle$, which were computed from the anticipatory spike trains.

trajectory is sufficiently smooth to allow the extrapolation $\theta(t+\tau) \approx \theta(t) + \omega(t)\tau$ so that HD cells can make a useful prediction of future HD values, reminiscent of predictive coding in sensory systems (Palmer et al., 2015). Random head movements with a frequency higher than $1/\tau$ would make it impossible to significantly reduce the bias by anticipation. Indeed, the measured movement correlations are in the range of seconds, much longer than the ATIs of HD cells. The importance of anticipation is underlined by the fact that the bias contributes as much to the error in estimating the HD as neural variability itself, at least for model populations of HD cells with experimentally measured parameters (Fig. 7).

Anticipatory firing might compensate for delays that inevitably arise in sensory transduction (Berry et al., 1999) or as information is processed by a multilayered system (Nijhawan and Wu, 2009), but this might be only part of anticipation's computational purpose. Compensating for delays, for example, does not explain why ATIs increase when rats are rotated passively (Bassett

et al., 2005), the delays across the HD system should remain unaffected. Indeed, the increase in ATIs also runs counter to the hypothesis that motor-efference signals cause the ATIs; passive rotation eliminates the motor efference and should cause the ATI to decrease, not increase. Yet a different experimental observation might have an important implication for decoding HD under passive rotation: firing rates in ADN decrease (Bassett et al., 2005). To counterbalance the resulting loss of information about HD, the read-out window T should increase (Burak and Fiete, 2012), and so should the ATI according to our framework. Likewise, HD cells with longer ATIs typically fire at a lower rate than cells with shorter ATIs (Blair et al., 1997). If downstream neurons sample from distinct subpopulations of HD cells, this systematic ATI difference would once again match the theoretical framework. Whether ATIs indeed adapt to maintain an optimal read-out is, as yet, unproven. Moreover, what appears to be adaptation, whether optimal or not, might simply be a byproduct of the biophysical mechanism that underlies the ATIs, much like

Figure 9. Anticipation improves the accuracy of the HD compass. **A**, Inset, The best accuracy $A(T)$ achievable for a homogeneous population with nonanticipatory neurons. The main figure represents the relative change of this smallest error if HD cells fire in an anticipatory manner. Data points are duration-weighted averages over all trajectory segments. Error bars indicate the duration-weighted SDs. For population sizes between 1000 and 10,000 neurons, the biologically realistic value $\tau = 25$ ms belongs to the best options and improves the accuracy by up to ~40%. Anticipation can, however, also deteriorate the code quality, in our examples, for $\tau = 50$ ms and population sizes $N > 6000$. **B**, Increasing the population size reduces the minimum error, too. The figure represents how much larger the population size N_0 of a nonanticipatory population needs to be to yield the same effect as an anticipatory population of size N (color code as in **A**). For example, the performance of a population with $N = 10,000$ neurons and an ATI of $\tau = 25$ ms matches that of a population with $N = 40,000$ without anticipation.

gain control in motion detection is intrinsic to the “Reichardt detector” mechanism (Borst et al., 2005).

If the anticipatory HD signal in the ADN were computed based on the read-out from another HD population, we expect even larger ATIs for that upstream population. As optimal ATIs change only slightly with system size (Fig. 9A), ATI differences of ~25 ms are likely to occur along the hierarchy of the HD system. The staggered ATIs found in the HD system (from $\tau = 40$ –75 ms in the rat LMN to $\tau \approx 25$ ms in the ADN and retrosplenial cortex to zero lag in the postsubiculum) (Taube, 2007) are consistent with this idea. In parallel, LMN activity could also be read out directly over a time window of 80–150 ms to yield a second estimate of current HD in addition to that provided by the ADN. A third and almost instantaneous estimate can be obtained from the postsubiculum. Consistent with our framework, the population size increases with decreasing ATI from LMN (~1000 cells) to ADN (~12,000 cells) to postsubiculum (~60,000 cells). Although these three subnetworks seem to face different trade-offs concerning network size and single-neuron dynamics, they may nevertheless provide excellent HD estimates for downstream computations in other brain areas. For instance, the grid-cell system in the entorhinal cortex may combine HD estimates based

on direct input from ADN and postsubiculum (Shibata, 1993; van Groen and Wyss, 1995) with distance information to achieve accurate path integration.

Furthermore, the hypothesis of optimal population-vector decoding leads to testable predictions for the ATIs in other species with different movement statistics, such as flies (Seelig and Jayaraman, 2015; Kim et al., 2017), cockroaches (Varga and Ritzmann, 2016), or bats (Finkelstein et al., 2015). All the information needed to calculate the optimal ATIs can be derived from the movement statistics, single-neuron properties, and network size. If the movements vary with the behavioral state, so should the ATIs. While cells in the dorsal subiculum of bats change the encoding of HD from pure to conjunctive coding when the animal switches from straight long-range navigation to rapid small-scale maneuvering (Finkelstein et al., 2018), the anticipatory behavior of these cells has not yet been addressed.

The HD system is a well-studied model, but it is only one particular example of how bias-variance trade-offs influence causal decoders of neuronal information. Computing error signals, as envisioned in theories of motor control (Wolpert et al., 1995; Bastian, 2006), also faces this trade-off. Neurons must gather sufficient information to reliably perform computations while still adapting quickly enough to reflect changing inputs throughout many hierarchical processing streams in the nervous system. Rather than being governed solely by biophysical limits on the speed of signal propagation, integration, and action-potential generation, the time constants of dendritic and somatic integration as well as axonal propagation delays might also reflect decoding constraints. Further work is needed to elucidate how the requirement that the nervous system operate in real time shapes processing delays.

References

- Barry C, Lever C, Hayman R, Hartley T, Burton S, O’Keefe J, Jeffery K, Burgess N (2006) The boundary vector cell model of place cell firing and spatial memory. *Rev Neurosci* 17:71–97.
- Bassett JP, Zugaro MB, Muir GM, Golob EJ, Muller RU, Taube JS (2005) Passive movements of the head do not abolish anticipatory firing properties of head direction cells. *J Neurophysiol* 93:1304–1316.
- Bastian AJ (2006) Learning to predict the future: the cerebellum adapts feedforward movement control. *Curr Opin Neurobiol* 16:645–649.
- Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. *Neuron* 76:695–711.
- Berry MJ 2nd, Brivanlou IH, Jordan TA, Meister M (1999) Anticipation of moving stimuli by the retina. *Nature* 398:334–338.
- Blair HT, Sharp PE (1995) Anticipatory head direction signals in anterior thalamus: evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *J Neurosci* 15:6260–6270.
- Blair HT, Lipscomb BW, Sharp PE (1997) Anticipatory time intervals of head-direction cells in the anterior thalamus of the rat: implications for path integration in the head-direction circuit. *J Neurophysiol* 78:145–159.
- Blair HT, Cho J, Sharp PE (1998) Role of the lateral mammillary nucleus in the rat head direction circuit: a combined single unit recording and lesion study. *Neuron* 21:1387–1397.
- Bobrowski O, Meir R, Eldar YC (2009) Bayesian filtering in spiking neural networks: noise, adaptation, and multisensory integration. *Neural Comput* 21:1277–1320.
- Borst A, Flanagan VL, Sompolinsky H (2005) Adaptation without parameter change: gain control in motion detection. *Proc Natl Acad Sci U S A* 102:6172–6176.
- Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18:7411–7425.
- Burak Y, Fiete IR (2012) Fundamental limits on persistent activity in networks of noisy neurons. *Proc Natl Acad Sci U S A* 109:17645–17650.
- Butler WN, Smith KS, van der Meer MA, Taube JS (2017) The head-

- direction signal plays a functional role as a neural compass during navigation. *Curr Biol* 27:1259–1267.
- Cheung A, Zhang S, Stricker C, Srinivasan MV (2008) Animal navigation: general properties of directed walks. *Biol Cybern* 99:197–217.
- Clark BJ, Sarma A, Taube JS (2009) Head direction cell instability in the anterior dorsal thalamus after lesions of the interpeduncular nucleus. *J Neurosci* 29:493–507.
- Clark BJ, Brown JE, Taube JS (2012) Head direction cell activity in the anterodorsal thalamus requires intact supragenual nuclei. *J Neurophysiol* 108:2767–2784.
- Finkelstein A, Derdikman D, Rubin A, Foerster JN, Las L, Ulanovsky N (2015) Three-dimensional head-direction coding in the bat brain. *Nature* 517:159–164.
- Finkelstein A, Ulanovsky N, Tsodyks M, Aljadeff J (2018) Optimal dynamic coding by mixed-dimensionality neurons in the head-direction system of bats. *Nat Commun* 9:3590.
- Georgopoulos AP, Caminiti R, Kalasja JF, Massey JT (1983) Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations. *Exp Brain Res* 7:327–336.
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 233:1416–1419.
- Glasius R, Komoda A, Gielen SC (1997) The population vector, an unbiased estimator for non-uniformly distributed neural maps. *Neural Netw* 10:1571–1582.
- Goodridge JP, Taube JS (1997) Interaction between the postsubiculum and anterior thalamus in the generation of head direction cell activity. *J Neurosci* 17:9315–9330.
- Hafting T, Fyhn M, Molden S, Moser MB, Moser EI (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436:801–806.
- Huys QJ, Zemel RS, Natarajan R, Dayan P (2007) Fast population coding. *Neural Comp* 19:404–441.
- Jammalamadaka SR, Sengupta A (2001) Topics in circular statistics. World Sci J. Singapore: World Scientific Press. Available from: <https://www.worldscientific.com/worldscibooks/10.1142/4031>.
- Johnson A, Seeland K, Redish AD (2005) Reconstruction of the postsubiculum head direction signal from neural ensembles. *Hippocampus* 15:86–96.
- Kim SS, Rouault H, Druckmann S, Jayaraman V (2017) Ring attractor dynamics in the *Drosophila* central brain. *Science* 356:849–853.
- Knierim JJ, Kudrimoti HS, McNaughton BL (1995) Place cells, head direction cells, and the learning of landmark stability. *J Neurosci* 15:1648–1659.
- Knierim JJ, Kudrimoti HS, McNaughton BL (1998) Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *J Neurophysiol* 80:425–446.
- Laurens J, Angelaki DE (2018) The brain compass: a perspective on how self-motion updates the head direction cell attractor. *Neuron* 97:275–289.
- Mardia K, Jupp P (2009) Directional statistics. New York: Wiley.
- Mathis A, Herz AV, Stemmler M (2012) Optimal population codes for space: grid cells outperform place cells. *Neural Comput* 24:2280–2317.
- Mizuseki K, Sirota A, Pastalkova E, Buzsáki G (2009) Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* 64:267–280.
- Mosheiff N, Agmon H, Moriel A, Burak Y (2017) An efficient coding theory for a dynamic trajectory predicts non-uniform allocation of entorhinal grid cells to modules. *PLOS Comput Biol* 13:e1005597.
- Mumford D (1992) On the computational architecture of the neocortex. *Biol Cybern* 66:241–251.
- Nijhawan R, Wu S (2009) Compensating time delays with neural predictions: are predictions sensory or motor? *Philos Trans A Math Phys Eng Sci* 367:1063–1078.
- O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34:171–175.
- Palmer SE, Marre O, Berry MJ 2nd, Bialek W (2015) Predictive information in a sensory population. *Proc Natl Acad Sci U S A* 112:6908–6913.
- Peyrache A, Lacroix MM, Petersen PC, Buzsáki G (2015) Internally organized mechanisms of the head direction sense. *Nat Neurosci* 18:569–575.
- Peyrache A, Schieferstein N, Buzsáki G (2017) Transformation of the head-direction signal into a spatial code. *Nat Commun* 8:1752.
- Puccini GD, Sanchez-Vives MV, Compte A (2007) Integrated mechanisms of anticipation and rate-of-change computations in cortical circuits. *PLoS Comput Biol* 3:e82.
- Rainer G, Rao SC, Miller EK (1999) Prospective coding for objects in primate prefrontal cortex. *J Neurosci* 19:5493–5505.
- Salinas E, Abbott LF (1994) Vector reconstruction from firing rates. *J Comp Neurosci* 1:89–107.
- Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser MB, Moser EI (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312:758–762.
- Schultheiss NW, Redish AD (2015) The compass within. *Nat Neurosci* 18:482–483.
- Seelig JD, Jayaraman V (2015) Neural dynamics for landmark orientation and angular path integration. *Nature* 521:186–191.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753.
- Shibata H (1993) Efferent projections from the anterior thalamic nuclei to the cingulate cortex in the rat. *J Comp Neurol* 330:533–542.
- Solstad T, Boccara CN, Kropff E, Moser MB, Moser EI (2008) Representation of geometric borders in the entorhinal cortex. *Science* 322:1865–1868.
- Stackman RW, Taube JS (1998) Firing properties of rat lateral mammillary single units: head direction, head pitch, and angular head velocity. *J Neurosci* 18:9020–9037.
- Stemmler M, Mathis A, Herz AV (2015) Connecting multiple spatial scales to decode the population activity of grid cells. *Sci Adv* 1:e1500816.
- Tan HM, Bassett JP, O'Keefe J, Cacucci F, Wills TJ (2015) The development of the head direction system before eye opening in the rat. *Curr Biol* 25:479–483.
- Taube JS (1995) Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J Neurosci* 15:70–86.
- Taube JS (2007) The head direction signal: origins and sensory-motor integration. *Annu Rev Neurosci* 30:181–207.
- Taube JS, Bassett JP (2003) Persistent neural activity in head direction cells. *Cereb Cortex* 13:1162–1172.
- Taube JS, Muller RU (1998) Comparisons of head direction cell activity in the postsubiculum and anterior thalamus of freely moving rats. *Hippocampus* 8:87–108.
- Taube JS, Muller RU, Ranck JB Jr (1990a) Head-direction cells recorded from the postsubiculum in freely moving rats: I. Description and quantitative analysis. *J Neurosci* 10:420–435.
- Taube JS, Muller RU, Ranck JB Jr (1990b) Head-direction cells recorded from the postsubiculum in freely moving rats: II. Effects of environmental manipulations. *J Neurosci* 10:436–447.
- Tsanov M, Chah E, Noor MS, Egan C, Reilly RB, Aggleton JP, Erichsen JT, Vann SD, O'Mara SM (2014) The irregular firing properties of thalamic head direction cells mediate turn-specific modulation of the directional tuning curve. *J Neurophysiol* 112:2316–2331.
- van der Meer MA, Knierim JJ, Yoganarasimha D, Wood ER, van Rossum MC (2007) Anticipation in the rodent head direction system can be explained by an interaction of head movements and vestibular firing properties. *J Neurophysiol* 98:1883–1897.
- Van Groen T, Wyss JM (1995) Projections from the anterodorsal and anteroventral nucleus of the thalamus to the limbic cortex in the rat. *J Comp Neurol* 358:584–604.
- Varga AG, Ritzmann RE (2016) Cellular basis of head direction and contextual cues in the insect brain. *Curr Biol* 26:1816–1828.
- Wiener SI, Taube JS (2005) Head direction cells and the neural mechanisms of spatial orientation. Cambridge, MA: Massachusetts Institute of Technology.
- Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. *Science* 269:1880–1882.
- Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* 16:2112–2126.
- Zhang K, Sejnowski T (1999) Neuronal tuning: to sharpen or to broaden? *Neural Comp* 11:75–84.
- Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ (1998) Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J Neurophysiol* 79:1017–1044.