J Neurophysiol 116: 1449–1467, 2016.
First published June 22, 2016; doi:10.1152/jn.00005.2016.

A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons

HyungGoo R. Kim,[1] Xaq Pitkow,[2,3] Dora E. Angelaki,[2,3] and Gregory C. DeAngelis[1]

[1]*Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York;* [2]*Department of Neuroscience, Baylor College of Medicine, Houston, Texas;* [3]*Department of Electrical and Computer Engineering, Rice University, Houston, Texas*

Submitted 4 January 2016; accepted in final form 16 June 2016

Kim HR, Pitkow X, Angelaki DE, DeAngelis GC. A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. *J Neurophysiol* 116: 1449–1467, 2016. First published June 22, 2016; doi:10.1152/jn.00005.2016.—Sensory input reflects events that occur in the environment, but multiple events may be confounded in sensory signals. For example, under many natural viewing conditions, retinal image motion reflects some combination of self-motion and movement of objects in the world. To estimate one stimulus event and ignore others, the brain can perform marginalization operations, but the neural bases of these operations are poorly understood. Using computational modeling, we examine how multisensory signals may be processed to estimate the direction of self-motion (i.e., heading) and to marginalize out effects of object motion. Multisensory neurons represent heading based on both visual and vestibular inputs and come in two basic types: "congruent" and "opposite" cells. Congruent cells have matched heading tuning for visual and vestibular cues and have been linked to perceptual benefits of cue integration during heading discrimination. Opposite cells have mismatched visual and vestibular heading preferences and are ill-suited for cue integration. We show that decoding a mixed population of congruent and opposite cells substantially reduces errors in heading estimation caused by object motion. In addition, we present a general formulation of an optimal linear decoding scheme that approximates marginalization and can be implemented biologically by simple reinforcement learning mechanisms. We also show that neural response correlations induced by task-irrelevant variables may greatly exceed intrinsic noise correlations. Overall, our findings suggest a general computational strategy by which neurons with mismatched tuning for two different sensory cues may be decoded to perform marginalization operations that dissociate possible causes of sensory inputs.

marginalization; multisensory; self-motion; heading; object

NEW & NOTEWORTHY

Sensory signals often reflect multiple variables that change in the environment. To estimate one task-relevant variable and ignore others, the brain needs to perform marginalization, which involves computing the probability distribution over one variable while averaging over irrelevant variables. We describe a computational approach by which a linear transformation of neural population responses can approximate marginalization. We show, through simulations involving diverse multisensory neurons, that this approach is effective in dissociating self-motion from object motion.

THE BRAIN ANALYZES INCOMING sensory signals to infer the structure of the environment and how it changes over time. In many cases, sensory signals reflect multiple stimulus variables that arise from different physical causes, yet only one of these causes may be of relevance to behavior in a particular context. For example, retinal image motion reflects both the motion of objects in the environment, as well as self-generated movements of the organism. If the subject wants to estimate their self-motion, then they need to ignore retinal image motion that is caused by object motion.

Optic flow provides a powerful visual cue for estimating the direction of self-motion, or heading (Gibson 1950). When an observer translates through a static environment without moving his head or eyes, the focus of expansion in the optic flow field indicates the observer's heading. Eye and head rotations distort the optic flow field and bias heading perception, but these distortions may be compensated by internal signals regarding eye and head movements (Crowell et al. 1998; Royden et al. 1992). Objects that move in the world also distort optic flow; however, there are no internal signals that can directly compensate for this distortion. Thus a fundamental question is how sensory signals are processed and decoded to estimate self-motion in the presence of object motion, or vice versa.

This task of dissociating self-motion from object motion can be framed as a marginalization problem (Beck et al. 2011; Pearl 2000), in which the brain computes the probability distribution over one stimulus variable while averaging over all others. Historically, these were called marginal probabilities, because in a table documenting the joint occurrences of two variables, the probabilities of either variable alone were written in the margin of the table as a sum over rows or columns. In the problem of dissociating self-motion and object motion, the response (r) of a visual neuron will generally depend on both heading (θ) and object direction (α): $r = f(\theta, \alpha)$. From such joint tuning profiles, it may be possible for the brain to compute the joint posterior probability distribution over both variables, $P(\theta,\alpha|\mathbf{r})$, where $\mathbf{r}$ represents the activity of a population of neurons. However, computing the joint posterior may not be practical, especially when there are several possible causes of sensory input (e.g., several moving objects). Instead, it is likely very useful for the brain to compute the marginal posterior over heading, for example, by integrating out the effects of object motion:

Address for reprint requests and other correspondence: G. C. DeAngelis, Dept. of Brain & Cognitive Sciences, Center for Visual Science, Univ. of Rochester, 310 Meliora Hall, Rochester, NY 14627-0268 (e-mail: gdeangelis@cvs.rochester.edu).

$$P(\theta|\mathbf{r}) = \int P(\theta, \alpha|\mathbf{r})d\alpha$$

While this operation is straightforward at the level of probability distributions, if neuronal activity represents probability distributions indirectly, then it may require nonlinear neural computations to implement optimally (Beck et al. 2011). Here, we explore linear decoding strategies that may provide an acceptable approximation to marginalization.

To what extent does the brain successfully dissociate self-motion and object motion? Psychophysical studies have shown that moving objects can bias visual heading perception (Dokka et al. 2015a; Fajen and Kim 2002; Layton and Fajen 2016; Royden and Hildreth 1996; Warren and Saunders 1995). Other studies suggest that observers can parse retinal image motion into components related to self-motion and object motion (Matsumiya and Ando 2009; Rushton et al. 2007; Warren and Rushton 2009a), but do not establish whether such parsing is partial or complete. Thus it remains unclear whether the visual system, by itself, can discount object motion to estimate heading. Vestibular inputs may also play important roles in dissociating self-motion and object motion, as they provide independent information about head translation (Angelaki and Cullen 2008) and can be integrated with optic flow for robust heading perception (Fetsch et al. 2009). Recent studies show that vestibular signals aid in compensating for self-motion when observers judge object motion (Dokka et al. 2015b; Fajen et al. 2013; Fajen and Matthis 2013; MacNeilage et al. 2012), and that vestibular signals reduce errors in heading perception in the presence of object motion (Dokka et al. 2015a). Thus, while vestibular signals clearly play a role in dissociating self-motion and object motion perceptually, the neural basis for this dissociation is unknown. We explore novel strategies for decoding responses of multisensory (visual/vestibular) neurons to marginalize over object motion and estimate heading.

Neurons with both visual and vestibular heading selectivity are found in the dorsal medial superior temporal area (MSTd; Duffy 1998; Gu et al. 2006), the ventral intraparietal area (VIP; Bremmer et al. 2002; Chen et al. 2011a, 2013; Schlack et al. 2002), and the visual posterior sylvian area (Chen et al. 2011b) of macaques. Some of these neurons have closely matched heading preferences for visual and vestibular stimuli ("congruent" cells), whereas others tend to prefer opposite headings for the two modalities ("opposite" cells) (Chen et al. 2011c; Gu et al. 2006). Congruent cells are well-suited to mediate cue integration (Chen et al. 2013; Gu et al. 2008) and cue weighting (Fetsch et al. 2012), whereas opposite cells are not because multisensory stimulation tends to flatten their heading tuning. Thus the functional role of opposite cells has remained mysterious. Interestingly, a neural representation involving a mixture of congruent and opposite cells has also been reported recently for coding of depth based on binocular disparity and motion parallax cues in the middle temporal (MT) area (Nadler et al. 2013), suggesting that such representations may have general utility that has not been appreciated thus far.

We demonstrate, through simulations, that decoding a mixed population of congruent and opposite cells, according to their vestibular heading preferences, can allow the brain to estimate heading more robustly in the presence of moving objects. Moreover, we show that this strategy is predicted by the optimal linear transformation of neural activity that marginalizes over object motion to estimate heading. Our findings suggest a general role for neurons with incongruent tuning preferences in performing marginalization operations, a computational strategy that may be useful across a range of sensory and motor systems. Importantly, our approach is very general and should apply broadly to marginalization problems involving other combinations of sensory signals and perceptual tasks.

METHODS

Definition of variables and description of presumed generative model. We begin by describing the general structure of the sensory input and then specify how we formulated the visual and vestibular responses of model neurons. We assume that the observer translates with heading θ and wants to estimate this direction. However, there may be an object in the visual field that moves in direction α. Both self-motion and object motion may influence the apparent visual stimulus direction θ_{vis} (as described in further detail below), whereas the vestibular system responds only to the direction of self-motion, $\theta_{\text{ves}} = \theta$. Neural responses $\mathbf{r}$ are driven by a combination of these two sensory variables, θ_{vis} and θ_{ves}. Two binary indicator variables, w_{vis} and w_{ves}, determine whether the visual and vestibular self-motion inputs are present, and a third binary variable w_{obj} indicates whether an object is present. Figure 6*A* depicts these dependencies as a probabilistic graphical model that defines the generative model that we assume for our model neural responses.

Simulation of heading tuning. We generated a population of 320 model neurons based on response properties of multisensory neurons that have been described previously in areas MSTd and VIP (Chen et al. 2011c; Gu et al. 2006). Each model neuron is selective for heading based on both visual (optic flow) and vestibular (head translation) inputs. The heading tuning curve for each sensory modality, in the absence of any object motion, is a von Mises function given by:

$$f_{\text{vis}}(\theta_{\text{vis}}) = A_{\text{vis}}\exp\{k_{\text{vis}}[\cos(\theta_{\text{vis}} - \theta_{\text{vis\_pref}}) - 1]\} + C_{\text{vis}} \qquad (1)$$
$$f_{\text{ves}}(\theta_{\text{ves}}) = A_{\text{ves}}\exp\{k_{\text{ves}}[\cos(\theta_{\text{ves}} - \theta_{\text{ves\_pref}}) - 1]\} + C_{\text{ves}}$$

where $f_{\text{vis}}(\theta_{\text{vis}})$ denotes the visual heading tuning function, $f_{\text{ves}}(\theta_{\text{ves}})$ denotes the vestibular tuning curve, and $\theta_{\text{vis\_pref}}$ and $\theta_{\text{ves\_pref}}$ indicate the visual and vestibular heading preferences, respectively. A_{vis} and A_{ves} represent tuning curve amplitudes, k_{vis} and k_{ves} control tuning widths, while C_{vis} and C_{ves} denote baseline response levels.

In some simulations, all model neurons were assumed to have identical response amplitudes and tuning widths (see Fig. 5*A*, tuning curve shape = constant). In these cases, the following parameters were used, although the results are robust to large variations in the choice of these parameters: $A_{\text{vis}} = A_{\text{ves}} = 50$, $k_{\text{vis}} = k_{\text{ves}} = 1$, $C_{\text{vis}} = C_{\text{ves}} = 5$. In other simulations, we created model neural populations with substantial diversity in the amplitudes and widths of tuning curves (see Fig. 5*A*, tuning curve shape = variable). In this case, parameters were drawn from uniform random distributions having the following ranges: for A_{vis} and A_{ves}, the range is [25, 75]; for k_{vis} and k_{ves}, the range is [0.7, 1.3]; and for C_{vis} and C_{ves}, the range is [0, 10]. For another set of simulations, to mimic the greater strength of visual responses typically seen in area MSTd relative to vestibular responses (Gu et al. 2006), we set the range of A_{ves} to be [12.5, 37.5] while keeping all other parameters as described above (see Fig. 5*A*, vestibular/visual strength: half).

We also considered variations in the model with respect to the distributions of visual and vestibular heading preferences. In the simplest version of the model, the heading preferences for each modality were created in exact steps of 45° (8 directions). We then created all combinations of visual and vestibular heading preferences, and there were five multimodal neurons for each combination ($8 \times 8 \times 5 = 320$; see Fig. 5*A*, preferred heading distribution = EqualStep). In this EqualStep configuration of the model, the five neurons having each combination of heading preferences had redundant tuning but independent variability; this was done to keep the number of neurons constant across model variants. We also created model populations involving two more realistic distributions of heading preferences. First, heading preferences of the 320 neurons were drawn randomly from a uniform distribution (0–360°) for each sensory modality (see Fig. 5*A*, preferred heading distribution = uniform). As a result, the difference in preferred heading between visual and vestibular tuning was also approximately uniformly distributed. Second, to mimic the bimodal distributions of visual and vestibular heading preferences that have been measured experimentally in areas MSTd and VIP (Chen et al. 2011c; Gu et al. 2006), we also simulated populations in which more neurons preferred one axis of motion (e.g., leftward or rightward) than the orthogonal axis (e.g., fore-aft, see Fig. 5*A*, preferred heading distribution = bimodal). Note, however, that our model formulation is general, and our results do not depend upon the particular axes of motion that are simulated.

Responses of model neurons to multisensory combinations of visual and vestibular stimuli (combined condition, f_{comb}) were modeled as a weighted linear sum of responses to unimodal stimuli, as justified by previous empirical studies (Fetsch et al. 2012; Morgan et al. 2008).

$$f_{\text{comb}}(\theta_{\text{vis}}, \theta_{\text{ves}}) = w_{\text{vis}} f_{\text{vis}}(\theta_{\text{vis}}) + w_{\text{ves}} f_{\text{ves}}(\theta_{\text{ves}}) \tag{2}$$

For simplicity, we assumed that responses to combined visual and vestibular inputs were simply the sum of the unimodal responses, but results are similar for other linear weightings. To compare with results from previous experimental studies (Chen et al. 2011c; Gu et al. 2006), we simulated responses for a combined condition in which both visual and vestibular inputs were active ($w_{\text{vis}} = w_{\text{ves}} = 1$), a visual condition in which heading was simulated by optic flow while there was no vestibular input ($w_{\text{vis}} = 1$, $w_{\text{ves}} = 0$), and a vestibular condition in which there is no visual motion but vestibular inputs signal self-motion ($w_{\text{vis}} = 0$, $w_{\text{ves}} = 1$).

Responses of each unit in individual trials were then generated from a Poisson distribution:

$$\mathbf{r} \sim \text{Poisson}\left[f_{\text{comb}}(\theta_{\text{vis}}, \theta_{\text{ves}})\right] \tag{3}$$

Real neural populations are likely to exhibit correlated noise, and very large populations of independent neurons would produce unrealistically confident estimates (Ecker et al. 2011; Shamir and Sompolinsky 2006). However, populations with realistic noise correlations should have similar properties as smaller uncorrelated populations (Gu et al. 2010; Moreno-Bote et al. 2014; Zohary et al. 1994). Since our primary aim is to examine the relative benefits of combining multimodal information on the biases induced by object motion, rather than to model the absolute precision of neural coding, our main results consider independent neuronal responses.

Interaction between background motion and object motion. Moving objects distort optic flow and can bias heading perception (Dokka et al. 2015a; Fajen and Kim 2002; Layton and Fajen 2016; Royden and Hildreth 1996; Warren and Saunders 1995). Thus, generally, the visual responses of model neurons are expected to reflect components of image motion associated with both self-motion and object motion. Unfortunately, little is known about how cortical neurons represent object motion and background motion associated with self-translation (Logan and Duffy 2006). More specifically, it is not clear from the literature how object motion alters the heading tuning of neurons. Given this dearth of knowledge, we did not attempt to model in detail how object motion and self-motion interact with the receptive field properties of neurons to determine neural responses. Rather, we assumed a worst-case scenario in which the visual responses of model neurons confound velocity components associated with self-motion and object motion. In other words, we assume that visual processing, on its own, cannot dissociate self-motion and object motion. If heading can be recovered by an appropriate decoding of multisensory neurons in this worst-case scenario, then the strategy should work for other (less severe) conflations of object motion and background motion. Importantly, because we simply combine object motion and self-motion into a single observable quantity, without modeling the details of visual stimuli, our formulation of the problem is broadly applicable to analogous problems involving other sensory systems and tasks.

Specifically, for each combination of heading (θ) and object direction (α), the visual stimulus direction, θ_{vis}, was assumed to be the vector average of velocity vectors associated with self-motion and object motion:

$$\theta_{\text{vis}} = \tan^{-1}\left(-\sin\theta + A_{\text{obj}}\sin\alpha, -\cos\theta + A_{\text{obj}}\cos\alpha\right) + \pi \tag{4}$$

where A_{obj} denotes the amplitude of object motion, and the amplitude of self-motion is assumed to be unity. When a moving object is not present in the simulations ($w_{\text{obj}} = 0$), $A_{\text{obj}} = 0$ and $\theta_{\text{vis}} = \theta$. The visual response of a model neuron was then computed according to *Eq. 1* and was combined with vestibular self-motion signals as indicated by *Eq. 2*.

Note that the visual tuning of each model neuron is solely dependent on the vector average of object motion and self-motion velocities. This causes estimates of heading derived solely from the visual responses of model neurons to be biased by objects that move in directions different from the optic flow caused by self-motion. By inherently confounding self-motion and object motion directions in the visual responses of model neurons, we provide a strict test of the hypothesis that a joint decoding of congruent and opposite cells can reduce biases induced by moving objects. Because of these properties, it is not possible for our model to solve the marginalization problem without taking advantage of a diverse population of multisensory responses. If it is possible to perform marginalization under this extreme case of neural encoding of visual signals, then it should be possible for a variety of other scenarios in which heading and object direction are less severely conflated in the visual responses of neurons.

We assumed a situation in which the speed of object motion was 1.5 times the speed of background motion (i.e., $A_{\text{obj}} = 1.5$), such that object and self-motion vectors did not cancel even when their directions were opposite. Readers may find it convenient to visualize self-motion and object motion as taking place within the frontoparallel plane (e.g., see Fig. 2*A*), such that optic flow patterns are simpler and all of the vectors lie in the vertical plane. However, since we simply treat self-motion and object motion as vectors, our formulation applies equally well to motion along other axes.

Population decoding by recognition model. To examine how well the brain could discount object motion α and estimate heading θ using a very simple decoder (not designed for marginalization), we used standard methods to approximate the likelihood function over heading from the population activity of model neurons. Assuming that neuronal activity follows independent Poisson statistics, that heading is drawn from a uniform prior, and that the neural tuning for heading is given by $f_i(\theta)$, the estimated log-likelihood $L(\theta)$ and log-posterior over heading $P(\theta|\mathbf{r})$ are given by (Dayan and Abbott 2001):

$$\begin{aligned}\log L(\theta) &= \log P(\theta|\mathbf{r}) + C_1 \\ &= \textstyle\sum_i r_i \log f_i(\theta) - \sum_i f_i(\theta) + C_2\end{aligned} \tag{5}$$

up to a stimulus-independent additive constant. It should be emphasized that the distribution obtained from *Eq. 5* will only equal the log-likelihood if a population of independent Poisson neurons represents heading and the shape of the population activity is not influenced by latent variables, such as object motion. If these conditions are not met (which will be the case in most of our simulations), then *Eq. 5* is not a true log-likelihood, but is instead called a recognition model (Hinton and Dayan 1996). We will therefore refer to this computation as a recognition model throughout the remainder of the paper.

From the distribution described by *Eq. 5*, we can obtain a heading estimate by computing the conditional expected heading (circular mean):

$$\hat{\theta} = \tan^{-1}\hat{\mathbf{z}} = \tan^{-1}\int \mathbf{z}P(\mathbf{z}|\mathbf{r})d\mathbf{z} \tag{6}$$

where $\mathbf{z}$ =(cos θ, sin θ) is a two-dimensional vector reflecting horizontal and vertical components of heading, $\hat{\mathbf{z}}$ is an estimate of this vector heading, and $\tan^{-1}\hat{\mathbf{z}}$ = extracts the heading direction (angle). The mean and standard deviation of heading errors were computed across 100 trials with different random samples of neural activity, to quantify the performance of the recognition model for each tested heading.

Since many of our model neurons have visual and vestibular tuning curves with different heading preferences, either of the two tuning curves could potentially be used in computation of the recognition model (*Eq. 5*). Indeed, one of our main results is that the choice of tuning curve, $f_i(\theta)$, greatly influences the heading estimation performance obtained from *Eq. 5* when moving objects are simulated. Thus we systematically compared the effects of decoding responses based on visual and vestibular tuning curves (see Fig. 4).

Although we do not expect this simple form of population decoding to perform marginalization, it provides an instructive comparison to an optimal linear decoder that does approximate marginalization, as described below.

Approximate linear marginalization. We sought to find a linear transformation of neural activity that can perform near-optimal marginalization; that is, one which computes an approximate marginal posterior distribution over heading that discounts the effects of moving objects. To perform what we call approximate linear marginalization (ALM; see text for details), we assume that the marginal posterior distribution is approximated by a member (Q) of the exponential family with linear sufficient statistics (Ma et al. 2006):

$$Q(\theta|\mathbf{r};\mathbf{h},g) = \frac{1}{Z}e^{\mathbf{h}(\theta)\cdot\mathbf{r}+g(\theta)} \tag{7}$$

We then optimize the parameters, $\mathbf{h}(\theta)$ and $g(\theta)$, to best explain data drawn from the true marginal posterior distribution, P. This optimization, performed using multinomial logistic regression (Bishop 2006), maximizes the likelihood of the parameters and minimizes the Kullback-Leibler divergence between the true marginal distribution and the approximation Q. Given K samples from the joint distribution (θ_k, α_k, $\mathbf{r}_k$) ~ $P(\theta,\alpha,\mathbf{r})$ for k = 1... K, the log-likelihood of the model parameters is given by

$$\log L(\mathbf{h},g) = \sum\nolimits_k \log Q(\theta_k|\mathbf{r}_k;\mathbf{h},g) = \sum\nolimits_k \left[\mathbf{h}(\theta_k)\cdot\mathbf{r}_k + g(\theta_k) - \log Z(\mathbf{r}_k,\mathbf{h},g)\right] \tag{8}$$

where $\mathbf{h}$ = [$h_i(\theta_k)$] denotes a matrix of weights specifying how the response r_i of neuron i influences the log-probability of heading θ_k on trial k; $g(\theta_k)$ represents a set of bias parameters for each heading; and the constant Z ensures that $Q(\theta_k|\mathbf{r}_k;\ \mathbf{h},\ g)$ is properly normalized (see main text for definition of Z). These parameters were obtained by multinomial logistic regression using the glmtrain() function in the Netlab toolbox (Bishop 2006; Nabney 2002), specifically with the algorithmic variant based on an approximate Hessian. We confirmed convergence by observing the likelihood values, changes of parameters, and cross-validation results for each iteration.

Note that this method is similar to the recognition model of *Eq. 5*, in that both use weighted sums of neural responses to construct a distribution over heading. However, whereas the weightings in *Eq. 5* were based on unimodal tuning curves (which generally do not provide the correct weightings for marginalization), the weightings in *Eq. 8* are learned directly from data to best approximate (minimize Kullback-Leibler divergence from) the true marginal posterior.

In our implementation of ALM, we allow 60 output nodes to represent different discrete possible values of heading, in steps of 6°. Each node j, corresponding to possible heading θ_j, receives input weighted by $\mathbf{h}(\theta_j)$ from all 320 model neurons in our multisensory population, as well as a bias $g(\theta_j)$. We generated a set of model population responses to train the ALM algorithm and tested its performance on a different set of model population responses. For individual trials of the training set, heading was selected randomly from one of 60 possible values spanning a range of 360° (6° resolution), and object direction was selected randomly from a uniform distribution having a range from 0 to 360°.

To account for the fact that there is generally variability and uncertainty regarding the sources of self-motion signals (visual, vestibular, or both), we constructed a training set that consists of a mixture of unisensory and multisensory self-motion experiences. Specifically, 10% of trials in the training set were vestibular-only self-motion conditions (e.g., an observer moving in the dark), 45% of trials were visual-only self-motion conditions for which there were no vestibular inputs (e.g., an observer driving at constant speed), and 45% of trials involved multisensory self-motion inputs. We chose these values somewhat arbitrarily based on the intuition that purely vestibular self-motion input (e.g., walking in the dark) is much less common than visual or combined self-motion input. However, our results are quite robust to considerable variation in these proportions, as long as there is at least a small portion (greater than ~20%) of training trials that include some vestibular input (vestibular-only or combined conditions).

We also explored how the performance of ALM depended on the prior probability of moving objects in the environment. To do this, we varied the probability that a moving object would be present in the training trials, and we examined how the decoding weights learned by ALM changed with object probability. Moving objects were only presented in the visual or combined self-motion conditions, because it was assumed that conditions leading to only vestibular self-motion input would have a low incidence of visible moving objects. If the probability of a moving object was 0.5, object motion would be simulated in one-half of the 45% (or 22.5%) of visual self-motion conditions and one-half of the 45% of combined self-motion conditions. We explored decoder performance for object appearance probabilities that ranged from zero (no moving objects in the training set) to one (all trials in the training set have moving objects except for the vestibular condition) in steps of 0.25. For each probability of object appearance, we generated neural responses corresponding to trial conditions having many combinations of heading and object motion drawn from the distributions described above. To train ALM using multinomial logistic regression, 10^6 training trials were presented for each probability of object appearance. Each training session of ALM required ~50 h of computing time on a 24-core computer. For each training set (involving a specific probability of object motion), the outcome of the fitting procedure was a set of decoding weights, $\mathbf{h}(\theta)$, for each of the 320 model neurons, and a set of bias terms, $g(\theta)$, for each possible heading.

Evaluation of cue-integration effects. Assuming that the likelihood functions computed for the single-cue conditions (visual and vestibular) can be approximated as independent Gaussian functions, we can use the standard formulation (Ernst and Banks 2002) to predict the optimal discrimination threshold for the combined condition from thresholds measured in response to visual and vestibular inputs separately:

$$T^2_{\text{optimal}} = \frac{T^2_{\text{vis}}T^2_{\text{ves}}}{T^2_{\text{vis}} + T^2_{\text{ves}}} \tag{9}$$

where T_{vis} and T_{ves} denote the heading discrimination thresholds derived from decoder output for the single-cue conditions, and T_{optimal} indicates the predicted heading discrimination threshold for the combined condition, assuming optimal cue integration of independent evidence. In addition, an empirical heading discrimination threshold for the combined condition, $T_{\text{empirical}}$, was also computed directly from the decoder output.

Reinforcement learning model. We sought to explore whether a simple network could learn to decode neural responses in a manner similar to ALM, when faced with the task of discriminating heading in the presence of moving objects. For this purpose, we used a reinforcement learning algorithm (see Law and Gold 2009 for details) to compute the connection weights that minimize errors in heading estimates. We built a simple two-layer network which performs a fine heading discrimination task. For visualization purposes (see Fig. 9*A*), self-motion direction was assumed to be upward with a small leftward or rightward component (±1, ±4, ±9, ±18°), and the network was trained to classify each stimulus as rightward or leftward. We simulated an object that could move in any random direction within the frontoparallel plane (0–360°) by adding a second visual velocity vector as described above for the main simulations. Note, again, that our formulation is general and would apply equally well to heading discrimination along any axis, as well as to other tasks and modalities. All combinations of heading and object directions were randomly interleaved in 200,000 training trials.

The network consisted of a population of 320 multisensory neurons in the first layer and a single readout neuron in the second layer. The multisensory input neuron population was constructed from bimodal distributions of heading preferences, as described above for ALM (equivalent to the bimodal H+V condition of Fig. 5*B*).

The response of the output (or decision) neuron, y, was computed as a weighted sum of the responses of the multisensory input neurons ($\mathbf{r}$), and decision noise (η) was added to the result:

$$y = \sum_i w_i r_i + \eta \tag{10}$$

Here, the vector $\mathbf{w}$ represents the weights of connections from the multisensory neurons to the decision neuron. Decision noise, η, was assumed to be normally distributed with zero mean and unit variance, $\eta \sim N(0, 1)$. If the output of the decision neuron (y) is positive, the network reports a rightward heading; if the result is negative, the network chooses leftward. A reward is given based on whether the choice is correct or not.

The weights in the network are updated based on a computation involving reward prediction error:

$$\Delta \mathbf{w} = \gamma C\left(R - \mathbb{E}\left[R\right]\right)\mathbf{r} \tag{11}$$

$$\mathbf{w}_{t+1} = \frac{\mathbf{w}_t + \Delta \mathbf{w}}{\|\mathbf{w}_t + \Delta \mathbf{w}\|_{\ell_2}} w_{\mathrm{amp}}$$

where γ ($=10^{-4}$) denotes the learning rate, C is the choice made by the network (-1 for a leftward heading decision, $+1$ for rightward), $\mathbb{E}[R]$ is the expected reward for a given decision, and R denotes whether the reward was received (0 for no reward and 1 for reward). To prevent weights from growing infinitely, weights are normalized to a constant ℓ_2 norm (Law and Gold 2009) of $w_{\mathrm{amp}} = 10$. Note that the weight change vector, $\Delta \mathbf{w}$, is proportional to the population response vector $\mathbf{r}$. If the choice is correct, then the pattern of weight changes is equal to the pattern of responses of the multisensory input neurons that gave rise to the choice. If the choice is incorrect, then weights change in the opposite direction to the input response pattern.

The expected reward function is given by (Law and Gold 2009):

$$\mathbb{E}\left[R\right] = \frac{1}{1 + e^{-|y|\beta}} \tag{12}$$

where y is the output of the decision unit from *Eq. 10* and β modulates the output. When $y\beta$ is near zero, the choice of the network is uncertain, and $\mathbb{E}[R]$ is close to 0.5 (chance). When $y\beta$ is a large negative or positive value, suggesting a more certain choice, $\mathbb{E}[R]$ is close to 1. The parameter β is critical because it controls the confidence in the network output and thus the size of weight change that results from a choice. Because the weights were initialized to random values, the value of y is initially small and unreliable, and β should be relatively large so that weight changes are sensitive to reward prediction error. As learning proceeds, the weights become optimized, and the values of y become larger and more reliable, such that β should decline with learning. We calculated β by using logistic regression to fit the relationship between the sign of stimulus heading (leftward vs. rightward) and y, using data from the most recent 100 trials.

Computing response correlations induced by object motion. We measured correlations in the responses of model neurons that were induced by the presence of a moving object in training trials (~10^6 trials). For a given object probability, we first computed the covariance between responses of pairs of model neurons for each heading (60 headings) and each stimulus modality (vestibular, visual, and combined conditions). Responses of model neurons were generated from an independent Poisson process, but since object direction varies from trial to trial, this introduces significant response covariation among model neurons, even for a single heading and stimulus modality. We then averaged the covariance matrix across headings and stimulus modalities to obtain a mean covariance matrix. Finally, the mean covariance matrix was transformed into a correlation matrix by dividing each element by the corresponding standard deviations.

In some simulations, described further below, we also added correlated noise to the model neural responses according to an empirically determined relationship between signal and noise correlations (Gu et al. 2011, 2014).

RESULTS

In simulations, we explored how object motion may bias estimation of self-motion, and how decoding a mixed population of congruent and opposite cells may allow more robust heading estimation. After describing the structure of the model, we explore how different approaches to decoding congruent and opposite cells affect heading estimation. Next, we consider the problem of estimating heading in the presence of moving objects as a marginalization problem, and we develop a near-optimal solution that can be achieved by a linear transformation of responses (which we term approximate linear marginalization, ALM). Finally, we explore how an effective decoding strategy may be learned autonomously using a simple reinforcement-learning model.

Model structure. We first consider how to model visual-vestibular integration of self-motion signals by single neurons and describe how we generated a model population of neurons ($N = 320$) with properties roughly similar to those observed in previous studies (Chen et al. 2011c; Gu et al. 2006). Based on previous theoretical (Ma et al. 2006) and empirical (Morgan et al. 2008) findings, we assumed that model neurons perform a weighted linear summation of their visual and vestibular inputs (*Eq. 2*) and generate responses independently according to a Poisson distribution with a mean rate given by the tuning curves. For simplicity, we assumed that neurons perform a straight sum of their inputs, even though empirical studies have demonstrated that weights tend to be subadditive (Chen et al. 2013; Fetsch et al. 2012; Morgan et al. 2008). However, our results are quite robust to this assumption (data not shown).

In the simplest form of the model, we assumed that visual and vestibular self-motion inputs have von Mises tuning curves with evenly spaced heading preferences (Fig. 1*A*) in steps of 45°. We then generated multisensory model neurons by drawing inputs from all possible combinations (8 × 8 = 64) of vestibular and visual heading preferences. The simplest neuronal population included five neurons for each heading preference combination, for a total of 320 model neurons (see METHODS for details). We first show results using this idealized population to build intuition and then discuss generalizations to

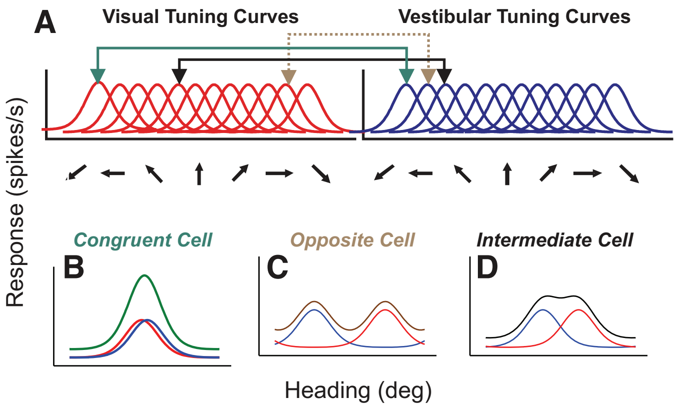

Fig. 1. Construction of model multisensory neurons. *A*: tuning curves of unimodal visual and vestibular input units, shown here with equally spaced heading preferences. Colored lines and arrows denote pairs of inputs that were combined to construct model neurons in *B–D*. *B*: to construct a congruent cell, visual and vestibular inputs with similar tuning preferences (green arrows in *A*) are combined. *C*: for a model opposite cell, visual and vestibular inputs with opposite tuning preferences (brown arrows in *A*) are combined. *D*: an intermediate cell, constructed from inputs with heading preferences that are neither congruent nor opposite (black arrows in *A*). In all cases, responses of the model neurons to combined visual and vestibular inputs are the sum of responses to the unimodal inputs (see METHODS).

the model that make the population tuning properties more realistic. When the two inputs to a multisensory model neuron have closely matched heading preferences, we call the resulting multisensory neuron a congruent cell (Fig. 1*B*). In this case, since the neuron sums its inputs, the tuning curve in response to combined visual and vestibular stimuli (Fig. 1*B*, green) has a single peak with an amplitude roughly double that of the unimodal tuning curves (Fig. 1*B*, red and blue). When the randomly drawn visual and vestibular inputs have very different heading preferences, we call the resulting multisensory neuron an opposite cell (Fig. 1*C*). In this case, the tuning curve for multisensory stimuli has two peaks centered near the visual and vestibular heading preferences of the inputs. Neurons with properties intermediate between congruent and opposite cells are also generated (Fig. 1*D*).

Having defined how model neurons integrate their visual and vestibular self-motion inputs, we now consider how object motion and self-motion interact to determine the visual responses of model neurons. When an observer sees an object that moves in the world (Fig. 2*A*), the retinal image motion of the object is the resultant of the observer's self-motion velocity and the velocity of the moving object in the world (Fig. 2*B*, black). Given that relatively little is known about how self-motion and object motion interact to determine the responses of cortical neurons (Logan and Duffy 2006), we did not attempt to model how the detailed receptive field properties of neurons would generate responses to background and object motion. Rather, we assumed a worst-case scenario in which our model neurons have no capability to distinguish between motion components related to self-motion and object motion. Specifically, we assume that the visual direction tuning of our model neurons is determined by the vector average direction of object motion and self-motion (Fig. 2*B*, white, see METHODS, *Eq. 4*). By modeling object motion and self-motion as generic vectors, our model is also quite general and can be applied readily to other sensory systems and tasks.

This approach ensures that, in the absence of vestibular inputs, the visual heading tuning of model neurons will be systematically biased by the presence of a moving object, both for congruent cells (Fig. 2*C*) and for opposite cells (Fig. 2*D*). The shifts in tuning caused by object motion are modest for congruent cells (Fig. 2*C*) because the vestibular and visual inputs are aligned to the true heading, causing a larger self-motion response that is modestly influenced by object motion. In contrast, the tuning curves of opposite cells (Fig. 2*D*) have substantially lower peak-trough amplitudes and are more greatly influenced by object motion, leading to larger peak shifts. Thus we start from a situation in which visual responses confound self-motion and object motion, such that the marginalization problem cannot be solved without taking advantage of diverse multisensory responses. We seek to determine whether an appropriate decoding of neurons with various visual-vestibular congruencies can recover accurate heading estimates in the presence of moving objects.

Population decoding by recognition models. To explore how responses of congruent and opposite cells might be combined to overcome the confounding of self-motion and object motion in visual responses, we used a well-established method (Dayan and Abbott 2001; Jazayeri and Movshon 2006; Ma et al. 2006) to compute the likelihood function over heading from the population response of conditionally independent model Pois-

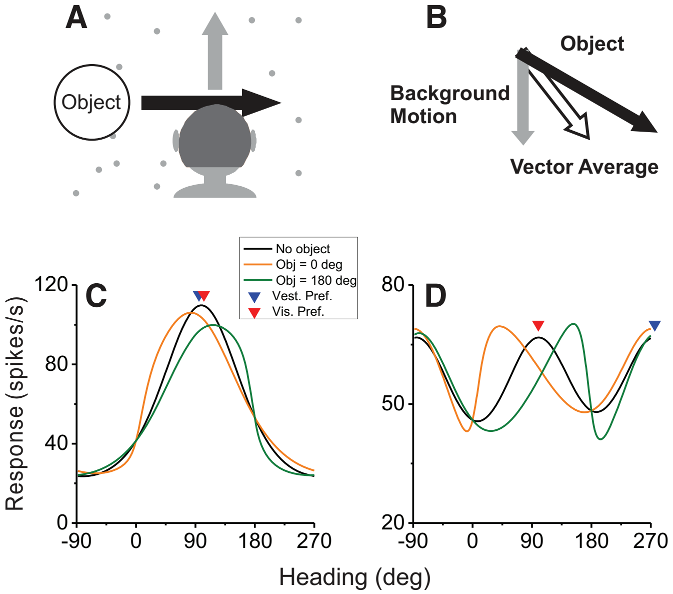

Fig. 2. Interaction between self-motion and object motion, and effects on model neurons. *A*: schematic illustration of a situation in which a large object moves rightward while the observer translates upward. Stationary components of the scene are depicted as dots in the background. *B*: retinal motion vectors corresponding to the situation depicted in *A*. Upward movement of the observer causes downward image motion of the background dots (gray arrow). Object motion on the retina (black arrow) is down to the right, reflecting the vector sum of self-motion and object motion in the world. Open arrow shows the vector average of the motion vectors corresponding to self-motion and object motion, and visual responses of model neurons were tuned as a function of this vector average direction to model a worst-case scenario (see METHODS). *C*: visual heading tuning of a model congruent cell in the absence of moving objects (black), and during simultaneous presentation of a rightward (0°, orange) or leftward (180°, green) moving object. Blue and red arrowheads denote the vestibular and visual heading preferences of the neuron, respectively. *D*: heading tuning curves for a model opposite cell, for the same conditions depicted in *C*.

son neurons with invariant heading tuning. Under these assumptions, the log-likelihood of each heading would be a weighted sum of the observed neural responses, with weights given by the logarithm of the mean response to each heading (*Eq. 5*). On the other hand, if the responses are not independent Poisson variables or there are other variables that alter the heading tuning on a trial-by-trial basis, then this computation would not produce the true log-likelihood but rather an estimate of the log-likelihood. The resultant distribution is sometimes called a "recognition model" to distinguish the model assumed for the purposes of recognizing stimuli from the true process that generated the sensory observations (Hinton and Dayan 1996). This recognition model is not designed to perform marginalization and is expected to return a biased estimate of the log-likelihood when other irrelevant variables, such as object motion, alter the shape of the population response (as occurs by design in our model population). We begin by exploring simple recognition models that use unimodal heading tuning curves as decoding weights, as these models build intuition and provide some novel insights into how diverse populations of multimodal neurons might be used to ignore irrelevant variables, such as object motion.

For multisensory neurons, the following question arises when estimating the log-likelihood from *Eq. 5*: which tuning curve should be used to compute the recognition model's log-likelihood? Two obvious possibilities are the visual tuning curve or the vestibular tuning curve (note that we do not consider the situation in which self-motion stimuli have discrepant visual and vestibular cues). For congruent neurons, the outcomes will be similar, since visual and vestibular tuning curves are almost identical. However, for incongruent neurons, including opposite cells, we consider whether it is more effective to compute the log-likelihood from the visual or vestibular tuning curves. Note that neither approach is guaranteed to be effective, and we consider a more general approach to obtaining decoding weights later. In addition, we explore the effects of selectively decoding all neurons, only congruent cells, or only opposite cells, to determine whether any of these simple decoding strategies could successfully marginalize over object motion to estimate heading.

We first consider an example case in which responses are decoded according to the vestibular tuning of each model neuron (Fig. 3). The log-likelihood over heading for the recognition model was computed for upward (90°) self-motion when there was no moving object (dashed curves) or when an object was moving rightward in the world (0°, solid curves). Figure 3*A* shows the log-likelihood function calculated from a sample of activity from congruent cells, defined as having a small difference in heading preference between visual and vestibular heading tuning curves ($|\Delta\text{preferred heading}| < 60°$). In the absence of a moving object (dotted curve), this function peaks very close to the true heading. However, when the object moves to the right, there is a small shift of the likelihood function toward leftward headings. In contrast, when the likelihood function is computed from the responses of opposite cells (defined as having $|\Delta\text{preferred heading}| > 120°$), object motion causes a moderate shift of the likelihood function toward rightward headings (Fig. 3*B*). Note that the shift induced by object motion is larger for opposite cells than for congruent cells, and that the peak-trough amplitude of the likelihood function is smaller for opposite cells. This pattern mirrors the effect of object motion on tuning curves (Fig. 2, *C* and *D*). Crucially, when the recognition model is computed from the activity of all neurons, we find that the bias induced by object motion is largely eliminated (Fig. 3*C*). Intuitively, this occurs because the opposite biases observed for congruent and opposite cells largely cancel each other when all neurons are decoded according to their vestibular heading preferences. The larger bias in the likelihood function for opposite cells (Fig. 3*B*) appears to be compensated by the greater amplitude of the likelihood function for congruent cells (Fig. 3*A*). This suggests a potentially simple multisensory decoding strategy to accomplish marginalization.

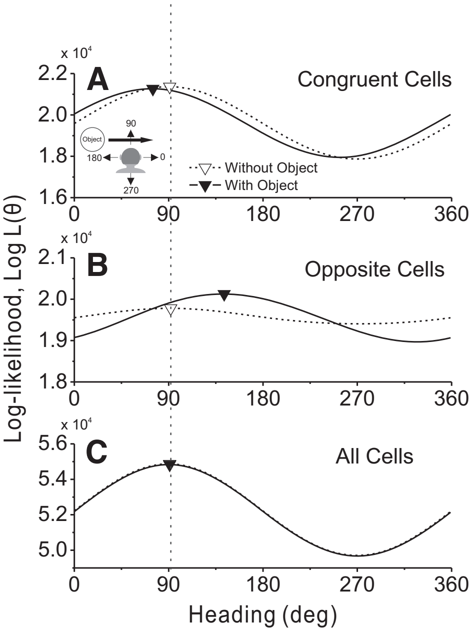

Fig. 3. Log-likelihood functions over heading computed from a single-trial sample of responses of the population of model neurons having equally spaced heading preferences and fixed tuning curves (EqualStep condition, see METHODS). In the situation depicted, self-motion is upward within the fronto-parallel plane (90°), and object motion is rightward (0°). Solid and dashed curves denote cases in which the moving object is present and absent, respectively. *A*: log-likelihood functions computed from congruent cells. The moving object biases the decoder's estimate of heading toward rightward. *B*: likelihood functions computed from responses of opposite cells. Object motion now biases the heading estimate in the opposite direction (leftward). *C*: when decoding all cells together, biases caused by object motion largely cancel and the heading estimate is accurate.

To summarize these observations, we computed the recognition model's log-likelihood function 200 times, with each simulated "trial" involving a new random sample of neural activity from the model neurons. Heading error was defined as the difference between the decoder's heading estimate (see METHODS, *Eq. 6*) and the true heading. We computed the mean and SD of heading errors (over the 200 simulated trials) for each of 12 different directions of object motion, with the self-motion direction fixed at 90° (upward). Figure 4*A* shows the results when each neuron's response is decoded according to its vestibular tuning curve. Decoding from congruent cells (open inverted triangles) produces substantial heading errors that depend systematically on the direction of object motion in the world. Similarly, decoding from opposite cells also yields large systematic errors (open circles), with the sign of the error being reversed relative to

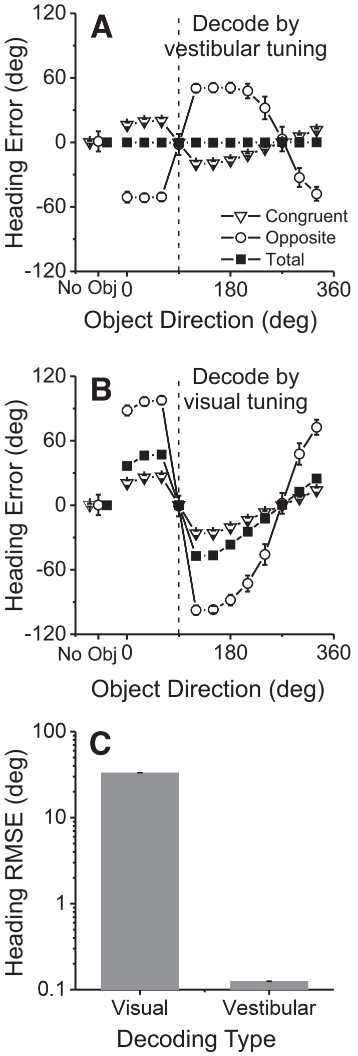

Fig. 4. Summary of heading errors for a model in which visual and vestibular inputs have tuning curves with equally spaced heading preferences and fixed tuning amplitude and width (EqualStep condition, see METHODS). *A*: heading error as a function of object direction when responses are decoded by computing the log-likelihood function using each neuron's vestibular tuning curve. Inverted triangles, circles, and squares denote decoding of congruent cells, opposite cells, or all cells, respectively. Data points labeled "No Obj" indicate heading errors when a moving object is not present. Data are shown for upward self-motion. *B*: heading error as a function of object direction when responses are decoded using each neuron's visual heading tuning curve. Format is the same as in *A*. *C*: root-mean-square error (RMSE) of heading estimates computed across all directions of self-motion and object motion. Decoding by vestibular tuning gives much smaller errors than decoding by visual tuning.

that obtained from congruent cells. Importantly, when we decode the responses of all neurons according to their vestibular tuning, heading estimation errors remain close to zero across the full range of object motion directions (solid squares in Fig. 4*A*).

Figure 4*B* demonstrates that this benefit of pooling responses across congruent and opposite cells does not occur when responses of model neurons are instead decoded based on their visual heading tuning curves. In this case, the errors that accompany decoding of congruent and opposite cells have the same sign, and there is no cancellation of biases when all neurons are decoded. Together, the results of Fig. 4, *A* and *B*, suggest that the pooled activity of congruent and opposite neurons may render heading estimation robust to object motion, but only if the recognition model's log-likelihood function is computed based on the vestibular tuning of each neuron.

The results of Fig. 4, *A* and *B*, are presented for a single direction of self-motion. To summarize the results for all combinations of self-motion and object motion directions (self-motion was varied in steps of 30° and object motion in steps of 10°), we calculated the root mean squared (RMS) error of heading estimates across all conditions. Figure 4*C* shows the RMS heading errors (mean ± SD obtained from 20 simulations) obtained when the activity of all model neurons (congruent and opposite cells together) was decoded based on either the visual or vestibular heading tuning of each neuron. Heading estimation errors are more than 100-fold smaller when responses are decoded according to the vestibular tuning curve, and this difference is highly significant ($P < 10^{-15}$, Wilcoxon rank-sum test).

Heading errors with more diverse model neurons. The results of Fig. 4 hinge on the fact that object motion produces opposite biases in congruent and opposite cells, and that these biases approximately cancel when the recognition model is computed from the multisensory responses of both groups of neurons. However, the simulations thus far are based on an idealized model population in which heading preferences are discrete, equal numbers of neurons prefer each discrete heading, and all neurons have the same amplitude and width of tuning. We, therefore, explored whether the strategy of decoding multisensory neurons according to their vestibular preferences is robust to greater diversity in the properties of the neural population.

We manipulated three aspects of tuning of the model neurons: distribution of heading preferences, diversity of tuning strength/shape, and balance of the two sensory modalities (see METHODS for details). First, in addition to the equally spaced heading preferences used above (Fig. 5*A*, EqualStep), we simulated two more realistic distributions of heading preferences: a uniform (but random) distribution (Fig. 5*A*, Uniform) and a bimodal distribution of preferences (Fig. 5*A*, Bimodal) similar to that observed for horizontal-plane heading tuning in areas MSTd (Gu et al. 2006, 2010) and VIP (Chen et al. 2011c, 2013). Second, in addition to simulations in which all neurons had identical tuning curve amplitudes and shapes (Fig. 5*A*, Constant), we also introduced substantial diversity in the amplitude, width, and baseline response of tuning curves within the model population (Fig. 5*A*, Variable, see METHODS for details). Third, to mimic the experimental observation that the amplitude of vestibular tuning curves of MSTd neurons is generally lower than the amplitude of visual tuning curves (Gu et al. 2007), we also simulated model populations in which the amplitude of vestibular responses is on average one-half of the visual response (Fig. 5*A*, Equal vs. Half; see METHODS). The model that combines all of these features (Bimodal H+V) is the most realistic approximation to real MSTd neurons. Simulations were performed based on model populations having all combinations of the three factors described above. Population responses were again decoded by computing a recognition model based on the vestibular tuning of each neuron, and RMS heading errors were averaged over 10 instantiations of each simulation, using an

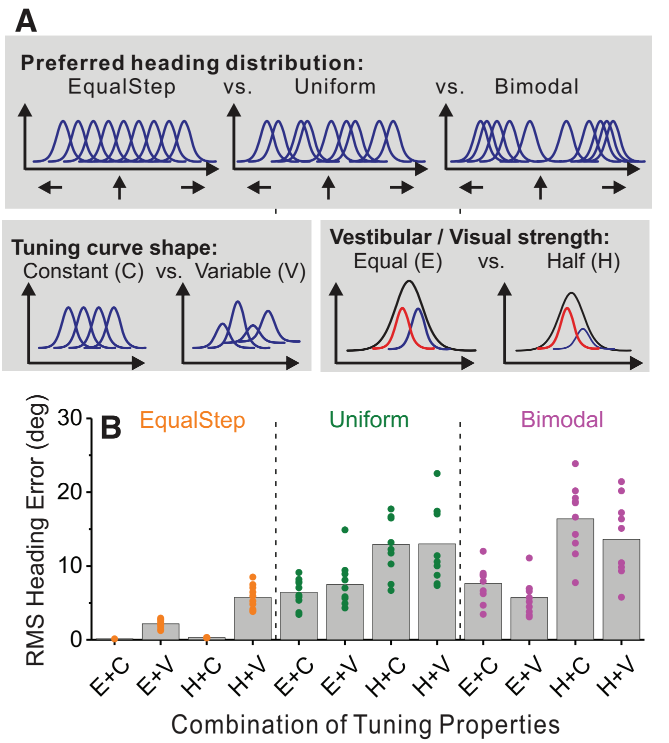

Fig. 5. Performance of the log-likelihood decoding model for more realistic properties of model neurons. *A*: illustration of the three aspects of input tuning curves that were varied. The distribution of heading preferences of input neurons could be equally spaced (EqualStep), Uniform, or Bimodal. The shape of tuning curves within the input populations was either constant (C) or variable (V) in amplitude, width, and baseline response. The strength of vestibular responses was either equal (E) or half (H) of the strength of visual responses. *B*: RMS heading errors for simulations corresponding to all permutations of the population tuning properties illustrated in *A*. Data are show for the case in which responses are decoded according to the vestibular tuning of each neuron.

independently generated neuronal population for each instantiation.

Figure 5*B* summarizes the RMS heading errors for each permutation of neuronal tuning properties. RMS heading errors increased substantially as the distribution of heading preferences changed from Equalstep to Uniform to Bimodal (main effect of distribution, three-way ANOVA, $P < 0.001$). Decreasing vestibular input strength (from Equal to Half) also caused a systematic increase in RMS heading error (main effect of vestibular strength, three-way ANOVA, $P < 0.001$). In contrast, variability in tuning curves did not yield a significant main effect on heading errors ($P = 0.84$), but had a significant interaction effect with heading preference distribution (three-way ANOVA, $P = 0.0013$). As a cumulative result of these effects, the most diverse model population (Bimodal H+V) shows much greater RMS heading errors (average RMS error = 13.6°) than the simplest model (EqualStep E+C). Performance in the Bimodal H+V condition is also substantially worse than that found in a recent behavioral study with animals (Dokka et al. 2015a); however, it should be kept in mind that such comparisons to behavior are severely limited by the fact that model performance depends on the number of neurons and their response properties. It is worth noting, however, that the most diverse population also produces substantially greater errors when only congruent cells (average RMS error = 27.7°) or only opposite cells (average RMS error = 134.8°) are decoded ($P = 0.002$ for both comparisons, Wilcoxon signed-rank test). Thus the benefit of decoding a mixed population of congruent and opposite cells is maintained, even though the absolute errors increase when more diverse properties are simulated. Results were very similar when larger neural populations were simulated ($N = 432$, 1,296, and 3,888, data not shown).

Together, these results suggest that the strategy of decoding multisensory neurons according to their vestibular heading preferences has some merit as an approach to marginalization, but it is not very robust to broad diversity in neuronal tuning properties.

Theory: near-optimal heading estimation by ALM. The results of Figs. 4 and 5 indicate that a recognition model based on vestibular tuning curves is partially successful in discounting the effects of object motion on heading estimates, provided that it takes advantage of a diverse population of multisensory neurons. This occurs despite the fact that the recognition model is not designed to perform marginalization. This led us to ask whether it is possible to determine a linear decoder of neural activity that can provide a good approximation to the true marginal posterior over heading. We present a general approach to this problem and compare the outcome to that of the recognition models described above.

Given that responses, **r**, of our model neurons are affected by both heading θ and object direction α, it is possible that the brain could estimate the joint posterior probability function, $P(\theta, \alpha|\mathbf{r})$, and thereby simultaneously estimate both heading and object direction. However, given that there could be multiple moving objects in a scene at one time, we do not know whether it is feasible for the brain to represent the joint posterior distribution over multiple such variables. Rather, in many situations, the observer may want to estimate one variable (e.g., heading) and simply ignore the other variable(s) (e.g., object motion). In this case, the optimal solution is for the brain to marginalize the joint posterior to eliminate the dependence on object direction:

$$P(\theta|\mathbf{r}) = \sum_{\alpha} P(\theta, \alpha|\mathbf{r}) \tag{13}$$

In general, the marginalized posterior may be a complicated function of **r**, even if the joint distribution $P(\theta,\alpha|\mathbf{r})$ is simple and easily decoded. For example, if the neural variability of the joint responses to heading and object motion falls into the exponential family with linear sufficient statistics (Ma et al. 2006), the joint posterior probability function can be easily computed by a linear decoder

$$P(\theta, \alpha|\mathbf{r}) = \frac{1}{Z} e^{\mathbf{h}(\theta,\alpha)\cdot\mathbf{r}+g(\theta,\alpha)} \tag{14}$$

where $\mathbf{h}(\theta,\alpha)$ is a matrix of weights used to transform the population response **r**, $g(\theta,\alpha)$ is influenced by the prior probability distribution over heading and object motion, and Z is a normalization constant defined below. Under these conditions, the marginalized posterior $P(\theta|\mathbf{r})$ will lie outside of the original exponential family of distributions with linear sufficient statistics and will generally require nonlinear transformations of neural responses for best performance (Beck et al. 2011).

We assume here that the brain is limited to processing neural responses linearly, even if this provides a suboptimal approximation of the desired quantities. This is equivalent to assuming that an approximation to the marginal posterior can be represented as a member of the exponential family with linear sufficient statistics,

$$Q(\theta|\mathbf{r};\mathbf{h},g) = \frac{1}{Z}e^{\mathbf{h}(\theta)\cdot\mathbf{r}+g(\theta)} \tag{15}$$

where Z is a normalization constant independent of θ,

$$Z(\mathbf{r};\mathbf{h},g) = \sum_{\theta} e^{\mathbf{h}(\theta)\cdot\mathbf{r}+g(\theta)} \tag{16}$$

We, therefore, attempted to find the parameters that best approximate the true marginalized posterior (P) by the form Q. Concretely, we sought the linear weighting functions $\mathbf{h}(\theta)$ and $g(\theta)$ that maximize the likelihood of the true heading over all responses. The quantity $\mathbf{h}(\theta)$ is a matrix in which each neuron (row) has a weight corresponding to each possible heading (column), whereas $g(\theta)$ represents a response-independent bias parameter for each heading. These desired quantities, $\mathbf{h}(\theta)$ and $g(\theta)$, can be computed using multinomial logistic regression (Bishop 2006). We refer to the log-linear model that best approximates the marginalized posterior as ALM.

While the above formulation focuses on marginalizing over the direction of object motion, the general inference problem faced by the brain in estimating heading is more complex. At any given time, information about self-motion can be provided by vestibular signals, visual signals, or both. In addition, moving objects are not always present in the scene. Thus estimating heading involves marginalizing over additional latent variables, in addition to object direction α:

$$P(\theta|\mathbf{r}) = \sum_{w_{\mathrm{obj}}=0}^{1}\sum_{w_{\mathrm{vis}}=0}^{1}\sum_{w_{\mathrm{ves}}=0}^{1}\sum_{\alpha} P\left(\theta,\alpha,w_{\mathrm{obj}},w_{\mathrm{vis}},w_{\mathrm{ves}}\middle|\mathbf{r}\right) \tag{17}$$

where w_{ves} is a latent variable that reflects the presence (or absence) of vestibular self-motion signals, w_{vis} reflects the presence or absence of visual self-motion signals, and w_{obj} indicates the presence or absence of a moving object. The assumed relationships among these variables are illustrated in the graphical model of Fig. 6*A*, which describes the generative model for the data that we have assumed.

In fact, we trained ALM to marginalize over these additional latent variables, in addition to object direction. As described in METHODS, we constructed a training set in which the presence of vestibular and visual self-motion signals was variable, and we trained ALM for a variety of probabilities of appearance of a moving object. Our findings are quite robust to the specific probabilities used to control the distributions over the latent variables for self-motion, as described in METHODS. Thus, while we focus on characterizing the performance of ALM in discounting the effects of object direction, our approach effectively marginalizes over multiple latent variables.

Robust heading estimation using ALM. We used ALM to estimate heading in the presence of moving objects. Specifically, we were interested in how well ALM could perform for the most diverse model neural population in Fig. 5 (Bimodal, H+V), which showed substantial heading errors when responses were decoded by computing log-likelihood according to the vestibular tuning of each neuron. For these simulations, model neural responses were generated in the same manner as described above (with tuning properties as in the Bimodal, H+V condition of Fig. 5), and the true headings were used to train ALM. Responses from 10^6 simulated trials were used to fit the data using multinomial logistic regression and to compute decoding weights for each neuron (see METHODS for details). Once the decoding weights were estimated for each neuron in the population, model performance was cross-validated by estimating heading from a new set of model neural responses (100 test trials for each combination of heading and object direction, 44,400 in total).

Figure 6*B* illustrates the decoder based on ALM in graphical form. There are 60 output nodes corresponding to all possible

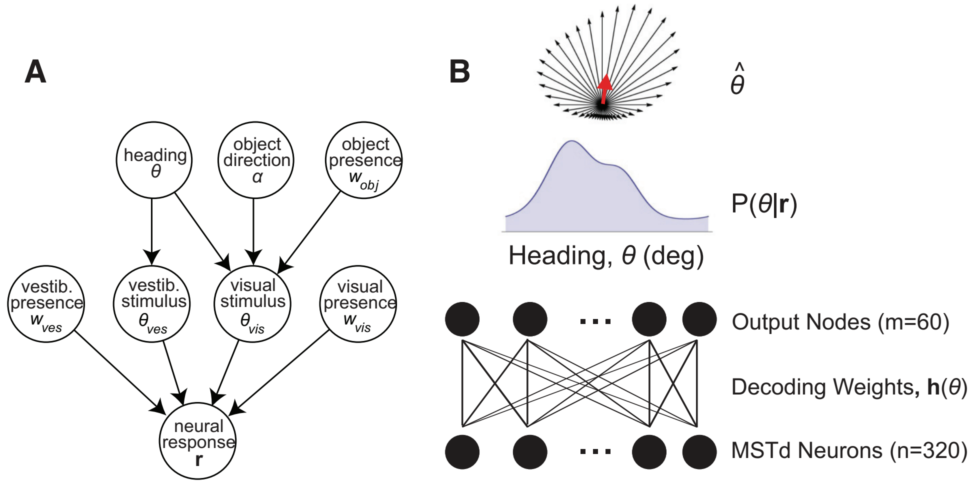

Fig. 6. Generative model for neural responses and characterization of an optimal linear decoder that performs approximate linear marginalization (ALM). *A*: graphical model illustrating the presumed generative model that gives rise to neural responses (**r**). The visual stimulus direction (θ_{vis}) depends upon both the observer's heading (θ) and the direction of object motion (α), whereas the vestibular stimulus direction (θ_{ves}) depends only upon heading (θ). Binary indicator variables determine the presence or absence of visual self-motion inputs (w_{vis}), vestibular self-motion inputs (w_{ves}), and object motion (w_{obj}). *B*: the ALM decoder takes input from a population of 320 model MSTd neurons. The decoder has 60 output nodes, each corresponding to a particular heading (in 6° increments). Each output node represents a weighted linear sum of the activity of the model MSTd neurons, with weights, $h_i(\theta)$, that are learned by multinomial logistic regression. The pattern of activity over the output nodes represents the log-posterior distribution over heading, log P(θ|**r**). The heading estimate, $\hat{\theta}$, is then computed as the circular mean of the activity of the output nodes (red arrow).

headings in 6° steps. Each output node represents a weighted linear sum of model MSTd responses, and the value at each output node represents the log probability that the stimulus contains the heading associated with that node. The approximate marginal posterior over heading was computed by exponentiating the output value of each node, and then normalizing such that the distribution over output nodes integrates to a value of one. For each simulated trial, heading was estimated as the circular mean (*Eq. 6*) of the marginal posterior distribution represented by the output nodes (red arrow in Fig. 6*B*).

To examine how the performance of ALM and the learned decoding weights depend on the environmental prevalence of object motion, we systematically varied the probability with which moving objects appeared in the training set. Zero probability indicates that there were no moving objects in the training trials that were used to fit the decoder weights, and a probability of one indicates that all training trials (except those with only vestibular self-motion) included a moving object, with random directions. We trained ALM separately for each different probability of appearance of moving objects. Importantly, after training, the RMS heading error was measured with a separate set of trials that all involved moving objects. Specifically, we quantified the error of the mean estimate over all trials with each given condition. Thus we examined how the probability of moving objects in the environment altered the strategy learned by the decoder.

Figure 7*A* (open squares) summarizes the RMS heading errors of ALM as a function of object probability in the training set. Note that all RMS errors shown in Fig. 7*A* are for cases in which moving objects were present (with random directions). For the combined condition, errors are quite large when the training set does not contain moving objects, as expected, but the error progressively decreases as the probability of object appearance increases (Fig. 7*A*, open black squares). For the visual condition (open red squares), performance of ALM is poor and independent of object probability because the visual responses of model neurons inherently confound self-motion and object motion. Importantly, performance of ALM in the combined condition (open black squares) far exceeds performance in the visual (open red squares) or vestibular (open blue squares) conditions when objects are common in the training set. This indicates that ALM takes advantage of both visual and vestibular signals to gain robustness to object motion.

Critically, as the object probability increases, performance of ALM becomes greatly superior to the performance obtained from a recognition model that computes likelihood functions from the vestibular or visual tuning of each neuron (Fig. 7*A*, circles and diamonds on the right margin). Note that we compared ALM and the recognition models on equal conditions because they were tested using the same diverse set of model responses (Bimodal H+V condition). Thus ALM copes substantially better with diversity in the multimodal tuning of the neural population than the recognition models.

To gauge the absolute level of performance of ALM, we also computed RMS heading errors by directly marginalizing the true joint posterior over heading, object direction, and the latent variables that describe the presence of self-motion inputs (w_{ves}, w_{vis}). For this purpose, we computed the joint tuning function of each model neuron for heading and object motion in each of the different self-motion conditions, $f_i(\theta, \alpha, w_{\text{ves}}, w_{\text{vis}})$, and then we calculated the joint posterior distribution under the assumption of independent Poisson neurons:

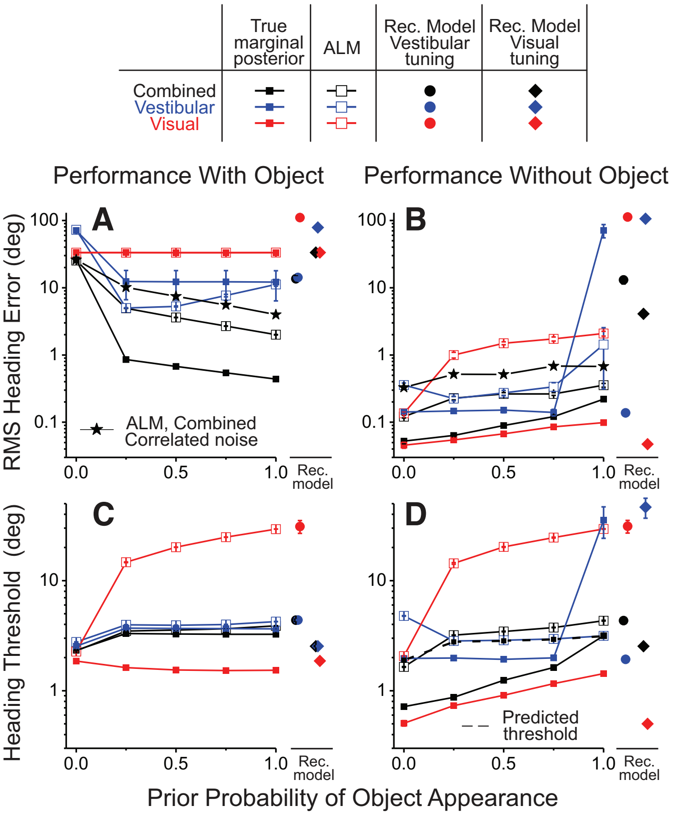

Fig. 7. Summary of performance of ALM, compared with the recognition models and the true marginal posterior. *A*: RMS heading errors, computed from test trials that always contained moving objects, are plotted as a function of the probability of appearance of a moving object. For each distinct stimulus condition, the error of the mean heading estimate is computed, and then the RMS value of this error is computed across all headings and object directions. Open squares show performance of ALM for the combined (black), vestibular (blue), and visual (red) conditions. Solid squares show performance from computation of the true marginalized posterior. Circles and diamonds show errors of the recognition models that involve decoding according to vestibular or visual heading tuning curves, respectively. Black stars show performance of ALM when empirically constrained correlated noise is added to model neuron responses (see text for details). *B*: RMS heading errors computed from test trials without moving objects. Format is the same as in *A*. *C*: heading discrimination thresholds for ALM, true marginal posterior, and recognition models. Thresholds shown here are for test trials that contain moving objects. Format is the same as in *A*. *D*: heading thresholds for test trials without moving objects. Format is the same as in *B*. Predicted thresholds for ALM, assuming optimal cue integration, are shown by the dashed black curve.

$$P(\mathbf{r} \mid \theta, \alpha, w_{\text{vis}}, w_{\text{ves}}) = \prod_i \frac{e^{-f_i(\theta,\alpha,w_{\text{vis}},w_{\text{ves}})} f_i(\theta, \alpha, w_{\text{vis}}, w_{\text{ves}})^{r_i}}{r_i!} \qquad (18)$$

$$P(\theta, \alpha, w_{\text{ves}}, w_{\text{vis}} \mid \mathbf{r}) \propto P(\mathbf{r} \mid \theta, \alpha, w_{\text{vis}}, w_{\text{ves}}) P(\theta) P(\alpha) P(w_{\text{vis}}) P(w_{\text{ves}})$$

We then integrated this joint posterior over α, w_{ves}, w_{vis}, and we estimated heading as the circular mean of the resulting distribution. This computation was performed for each stimulus condition, matching both the proportions of different trial types and modalities in the training set (see METHODS) and the object probability used in training ALM.

Performance of ALM relative to the true marginal posterior can be examined by comparing the open and solid squares in Fig. 7*A*. For the combined condition, ALM does not perform as well as the true marginal posterior (open vs. solid black squares), which is expected, given that optimal marginalization of neural responses with Poisson variability generally requires nonlinear operations (Beck et al. 2011), whereas ALM is a linear approximation to marginalization. At large object probabilities, performance of ALM is comparable to, if not slightly better than, performance of animals in a recent study of heading discrimination in the presence of object motion (Dokka et al. 2015a), whereas performance based on the true marginal posterior is substantially better than behavior (with the caveat noted above). For the visual condition, performance of the true marginal posterior is poor and identical to that of ALM (solid vs. open red squares in Fig. 7*A*), as expected since there is no way to disambiguate self-motion from object motion when only the confounded visual cues are provided. For the vestibular condition, performance of ALM is slightly better than that of the true marginal posterior (open vs. solid blue squares, Fig. 7*A*). This suggests that ALM has traded off some performance in the combined condition for better performance in the vestibular condition. Together, the results of Fig. 7*A* show that ALM achieves substantial robustness to object motion while retaining the simplicity of a linear transformation of neural activity. Moreover, convergence of visual and vestibular self-motion signals, combined with the existence of both congruent and opposite cells, is critical for achieving a robust approximation to the marginal posterior over heading.

Does ALM gain robustness to object motion at the expense of performance when moving objects are not present? To address this question, we also computed RMS heading errors for both ALM and the true marginal posterior for test conditions in which only self-motion was present and there was no moving object. As shown in Fig. 7*B*, heading errors are generally quite small in the absence of object motion. In the combined condition, both ALM and the true marginal posterior produce very small errors (open and solid black squares). ALM also produces very small errors in the vestibular condition (open blue squares), largely similar to the true marginal (solid blue squares). However, when the true marginal is computed under the assumption that object probability = 1, the true marginal produces large errors in the vestibular condition. This occurs because the responses of model neurons in the vestibular condition without objects are roughly threefold smaller than the response of model neurons in most other conditions, for which there are both visual and vestibular components of response. This makes the likelihood, $P(\mathbf{r}|\theta, \alpha, w_{\text{vis}}, w_{\text{ves}})$, of observing test trials with similarly low responses extremely low. In the visual condition, heading errors are substantially larger for ALM (open red squares) than the true marginal, but still comparable to behavioral heading thresholds (Dokka et al. 2015a). This reveals that ALM compromises a bit of performance in the absence of object motion to gain robustness to object motion.

Sensitivity of heading estimates based on ALM. Thus far, we have examined biases in heading estimates produced by ALM; however, a more complete evaluation necessitates examining the sensitivity of performance as well. This is quantified by computing heading thresholds, which are proportional to the standard deviation of the heading estimate. Figure 7*C* summarizes heading thresholds in the same format as Fig. 7*A*. For the vestibular (blue) and combined (black) conditions, heading thresholds exhibited by ALM are small and very similar to those computed from the true marginal posterior. For the visual condition, heading thresholds are much higher for ALM (open red squares) than for the true marginal (solid red squares), which is expected due to the confounded nature of the visual responses of model neurons. In the absence of object motion (Fig. 7*D*), heading thresholds are similar overall to the case of moving objects (Fig. 7*C*). The only notable difference is that vestibular thresholds of the true marginal are very large for the case of object probability = 1, the same case for which large heading errors were observed in Fig. 7*B*. Overall, these comparisons reveal that the sensitivity of heading estimates in the combined and vestibular conditions is only modestly reduced compared with true marginalization of the joint posterior.

Optimality of cue integration using ALM. Whereas our results show that decoding a mixed population of congruent and opposite cells can provide robustness to object motion, previous work has shown that decoding both congruent and opposite cells does not allow for optimal cue integration of visual and vestibular heading signals in the absence of object motion. Specifically, selective decoding of congruent cells could predict optimal cue integration, but inclusion of opposite cells led to grossly suboptimal performance (Fetsch et al. 2012; Gu et al. 2008). This suggests that there may be a tradeoff between cue integration and discounting of object motion.

To examine this issue, we compared heading thresholds obtained by ALM for the visual and vestibular conditions with thresholds obtained in the combined condition (open squares, Fig. 7*D*). If the evidence from each cue were independent, then the threshold for cue integration in the combined condition could be computed from the single-cue thresholds (*Eq. 9*). When ALM is trained to estimate heading in the absence of moving objects (zero object probability), the threshold obtained by ALM for the combined condition is very close to this prediction (Fig. 7*D*, dashed black curve). As the probability of moving objects in the training set increases, the combined threshold from ALM also increases with object probability and exceeds the optimal predicted threshold (Fig. 7*D*, open black squares vs. dashed black curve). Thus ALM does not perform optimal cue integration for heading discrimination when trained in the presence of moving objects, presumably because the weights are learned to reduce biases induced by moving objects and are no longer optimal for performing cue integration in the absence of moving objects. However, the loss of sensitivity for cue integration is modest: for an object appearance probability of one, the empirical combined threshold (4.37°) exceeds the optimal predicted threshold (3.15°) by 39%. For an object appearance probability of 0.5, the empirical threshold exceeds the optimal prediction by only 21%. Thus it is possible to learn a linear readout that provides considerable robustness to object motion with only a modest loss of sensitivity for cue integration.

Weight profiles learned by ALM. We have shown that a heuristic strategy (a recognition model that decodes a likelihood generated according to the vestibular tuning) and ALM both take advantage of vestibular signals to reduce errors in heading estimation caused by object motion. An interesting question is whether the decoding strategy learned by ALM bears some resemblance to the heuristic strategy of decoding by vestibular tuning.

To investigate this issue, we plotted the decoding weights learned by ALM for all 320 model neurons (Fig. 8, *A*–*D*). Each row in these matrices gives the decoding weights, $h_i(\theta)$, for the *i*th neuron in the model population. Weight profiles are sorted (from top to bottom) by each neuron's vestibular heading preference (left column) or its visual heading preference (right column). When there were no moving objects in the set of training trials that was used to determine the decoding weights, we observed a clear diagonal pattern (Fig. 8, *A* and *B*) in the weight profiles, indicating that the peak of the learned weight profiles had a tendency to match the vestibular or visual heading preferences of the neurons. For example, a model cell that prefers rightward motion from vestibular cues will generally contribute most to the output node of the decoder that represents rightward motion. This diagonal pattern is more clearly visible in Fig. 8*B*, relative to Fig. 8*A*, because the amplitude of the visual heading tuning curves of the model neurons was double that for the vestibular tuning curves.

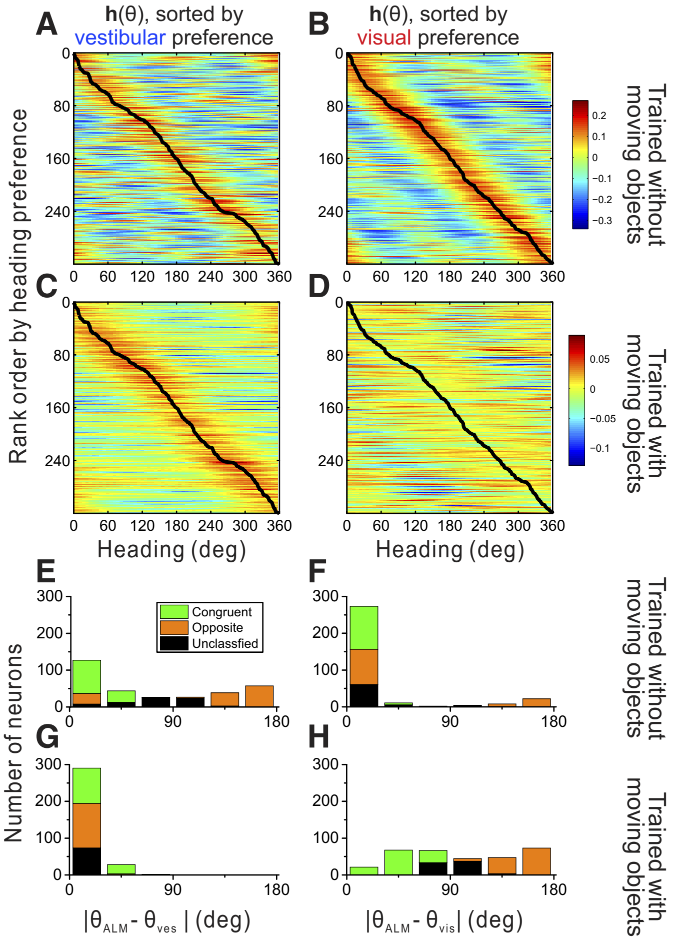

Fig. 8. Summary of decoding weights learned by ALM. *A*: decoding weights learned by ALM from a training set that did not contain moving objects. Each row in the matrix shows the decoding weights (color coded) for one of the 320 model MSTd units. Data are ordered from *top* to *bottom*, according to the vestibular heading preference of each unit. The black line connects the vestibular heading preferences for all of the neurons. The diagonal band structure indicates that the peak of a unit's decoding weight profile tends to match its vestibular heading preference. Biases [$g(\theta)$, see METHODS] are not shown for visualization purposes. *B*: decoding weights from *A*, now sorted according to the visual heading preference of each unit (visual preferences are connected by the black line). *C*: decoding weights learned by ALM from a training set that contained moving objects. Weight profiles are sorted according to vestibular heading preferences. Format is the same as in *A*. *D*: decoding weights from *C*, now sorted according to visual heading preferences. Format is the same as in *B*. *E*: histogram of the absolute difference between the vestibular heading preference of each neuron (θ_{ves}) and the heading at which its weight profile peaks (θ_{ALM}), for the case of training without moving objects. *F*: histogram of the absolute difference between the visual heading preference of each neuron (θ_{vis}) and θ_{ALM}, for the case of training without moving objects. *G*: same as *E*, but for the case in which ALM is trained with moving objects. *H*: same as *F*, but for the case in which ALM is trained with moving objects.

Note that the peak of the weight profile cannot align well to the preferred heading for all neurons because the population includes both congruent and opposite cells, and the decoding weights cannot align with both heading preferences for opposite cells. This causes the secondary weak diagonal bands, offset by 180°, which are visible in Fig. 8*B* (and to a lesser extent in Fig. 8*A*). These results, for the case of training without objects, are summarized in the histograms of Fig. 8, *E* and *F*. Here, we calculated the difference between the heading direction at which the ALM weight profile peaks (θ_{ALM}) and the vestibular (θ_{ves}) or visual (θ_{vis}) heading preferences of each model neuron. It can be seen that peaks of the weight profiles tend to align with visual and vestibular heading preferences, especially for congruent cells. The alignment is stronger with visual heading preferences (values closer to zero in Fig. 8*F* than 8*E*), simply because visual heading tuning was stronger than vestibular heading tuning in the model population used with ALM.

In contrast, when all of the trials in the training set include moving objects, a different pattern of results is obtained (Fig. 8, *C* and *D*). The peak of the weight profiles is now generally well matched to the vestibular heading preference of each neuron (diagonal band in Fig. 8*C*), but is not closely related to the visual heading preferences of the neurons (more diffuse pattern in Fig. 8*D*). Thus, despite the model neurons having weaker vestibular tuning than visual tuning, ALM learns to decode responses roughly according to the vestibular heading preference when moving objects are present in the training set. Figure 8*G* shows that the vestibular heading preference matches closely with the peak of the learned weight profile, with the vast majority of neurons showing differences <30°. In contrast, visual heading preferences are not well matched to θ_{ALM} (Fig. 8*H*) when ALM is trained with moving objects.

These results show that the heuristic strategy of decoding responses according to their vestibular tuning, as suggested previously based on choice probability data (Gu et al. 2014, 2008), roughly approximates an optimal linear solution for marginalizing over object direction to compute heading. However, it should be noted that the weights learned by ALM are not identical to vestibular tuning, and these differences allow ALM to better cope with diversity in the neural response properties.

Reinforcement learning of a near-optimal decoding strategy. Is it plausible that the brain could learn to decode activity from multisensory neurons in the correct manner to implement ALM? To explore this question, we constructed a simple decision-making network to perform a fine heading discrimination task (with and without moving objects, Fig. 9*A*), and we trained the network using a standard reinforcement learning algorithm (see METHODS; Law and Gold 2009; Sutton and Barto 1998), which is thought to be biologically plausible.

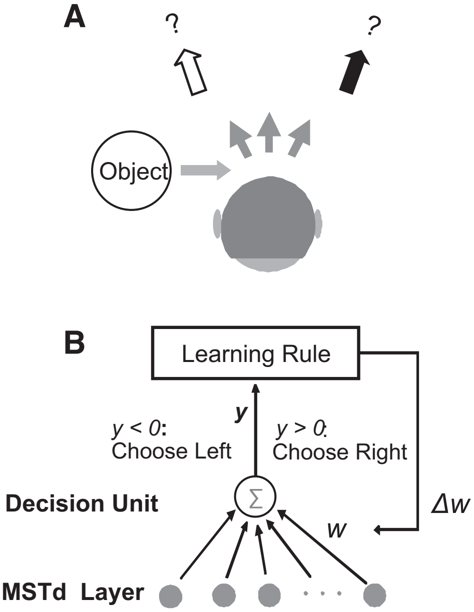

Fig. 9. Reinforcement learning network. *A*: the network was trained to perform a fine discrimination task. Based on the activity of a population of model MSTd neurons, the network is trained to determine whether self-motion is leftward or rightward of vertical, irrespective of any object motion. Objects moved in random directions across simulated trials, and the network was required to discount object motion to perform optimally. *B*: the network is composed of a layer of MSTd input units and a decision unit. The input units have tuning properties corresponding to the bimodal H+V condition (Fig. 5*B*). The decision unit performs a weighted linear sum of its inputs, to compute a decision variable, *y*. If $y > 0$, the network chooses rightward; if $y < 0$, it chooses leftward. Reinforcement learning was used to update the decoding weights, *w*, based on whether the choice was correct or not (see METHODS for details).

Figure 9*B* illustrates the structure of the network. It consists of an array of 320 model neurons that were constructed in the same manner as described above for the ALM analysis (with tuning properties as in the Bimodal, H+V condition of Fig. 5; results were very similar for other model populations with uniform heading preferences). This population of model neurons feeds into a single "readout" neuron, which performs a weighted linear sum of its inputs, with weights that are learned by the reinforcement learning algorithm. If the weighted sum of responses is greater than zero, the network chooses rightward; otherwise, it chooses leftward. Weights are updated according to whether the "decision" of the network is correct or not (see METHODS; Law and Gold 2009).

In one set of simulations, we trained the network to discriminate heading in the absence of moving objects. In this case, the readout weights depend systematically on both the vestibular and visual heading preferences of the model neurons (Fig. 10, *A* and *B*). Specifically, neurons with heading preferences to the right of vertical (90°) tend to show positive weights, and those with preferences to the left of vertical tend to show negative weights. Note, however, that opposite cells cannot obey this relationship for both visual and vestibular preferences, thus adding variability to the weight profiles. The relationship between readout weights and vestibular heading preferences (Fig. 10*A*) is more variable than the relationship with visual heading preferences (Fig. 10*B*). This is because visual tuning curves have double the amplitude of vestibular tuning curves, and the decoder gives greater weight to the stronger visual responses, such that opposite or intermediate cells have less systematic correspondence between vestibular preferences and decoding weights.

In a second set of simulations, we again trained the network to discriminate heading around straight forward, but each simulated trial also included an object that moved in a random direction. As described above, the moving object biased the visual responses of model neurons. Under these conditions, a very different pattern of weights was obtained by reinforcement learning. Specifically, the readout weights depended strongly on the vestibular heading preferences of the model neurons (Fig. 10*C*), but showed little systematic dependence on visual heading preferences (Fig. 10*D*) because the random biases from the object motion caused visually driven signals to be less reliable. In other words, when faced with the task of discriminating heading in the presence of moving objects, the network learned to decode responses according to the vestibular preference of the model neurons.

How does the decoding strategy attained through reinforcement learning compare with the weights learned by ALM? To make this comparison, we applied a variant of ALM to the heading discrimination task (Fig. 9). Since the task involves a binary choice, ALM was trained to classify heading as leftward or rightward based on neural responses, and we used binomial logistic regression to learn the decoding weights. This approach is very similar to the approach from the previous sections, except that it only computes the probability ratio for two nearby headings instead of computing the probabilities for all possible headings. Two hundred thousand trials were used to train ALM and the reinforcement learning algorithm, such that we could directly compare the decoding weights of model neurons for the two algorithms. Figure 11*A* shows *z*-scored readout weights as a function of vestibular heading preference

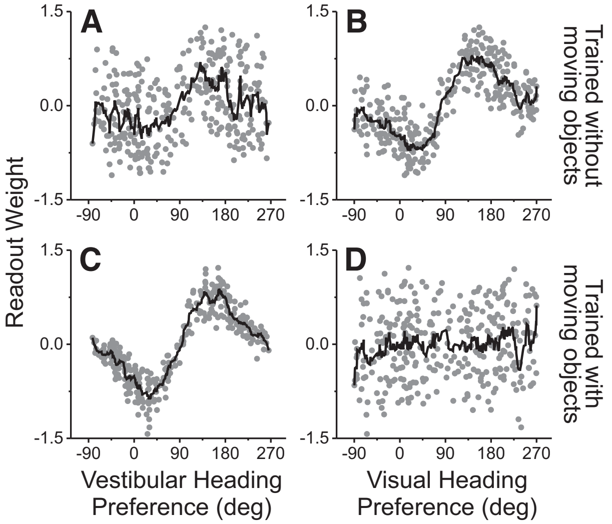

Fig. 10. Summary of readout weights obtained by the reinforcement learning algorithm. Each panel shows the readout weight for each model MSTd neuron plotted as a function of the neuron's vestibular (*A* and *C*) or visual heading preference (*B* and *D*) (gray symbols), along with the 10 point moving average (black line). *A* and *B*: decoding weights learned when the training set does not include any moving objects. Weights have a systematic dependence on both vestibular and visual heading preferences. *C* and *D*: decoding weights obtained by reinforcement learning when the training set includes moving objects. In this case, weights depend systematically on the vestibular heading preferences of the model MSTd units, but not on their visual heading preferences.

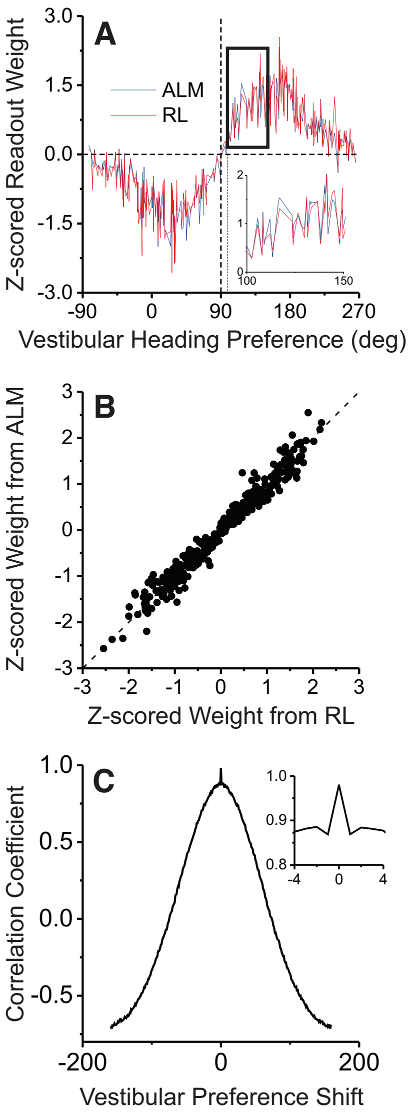

Fig. 11. Comparison of decoding weights obtained from ALM and from the reinforcement learning (RL) algorithm. Both algorithms were trained to perform fine heading discrimination based on the same set of model MSTd inputs. *A*: decoding weights are plotted as a function of the vestibular heading preference of each model MSTd unit. Data are show for ALM (blue) and for reinforcement learning (red). Weights for each model are normalized by *z*-scoring so that they can be directly compared. *Inset*: magnification of weight profiles within a narrow range. *B*: scatter plot of *z*-scored weights obtained from ALM vs. weights obtained from RL. Each datum represents weights for 1 of the 360 model MSTd input units. *C*: cross-correlogram of the two weight profiles shown in *A*. *Inset*: magnification of the central narrow peak of the correlogram.

for the two learning methods. Notably, weights obtained by reinforcement learning are remarkably similar to the weights obtained by ALM, and the two are highly correlated (Fig. 11*B*, $R = 0.98$, $P < 10^{-15}$, Pearson correlation). Even small fluctuations in weights among neurons with similar heading preferences are consistent between the two learning mechanisms (Fig. 11*A*, *inset*). This fine scale agreement between the two sets of weights is also revealed by computing a cross-correlation function between the two sets of weights. Figure 11*C* shows that shifting the alignment of the weights by one neuron (where neurons are ordered by their vestibular heading preferences) decreases the correlation coefficient from 0.98 to 0.83. These results show that a biologically plausible reinforcement learning algorithm can recover decoding weights that are nearly identical to those of the optimal linear decoder for approximating the marginal posterior over heading.

Correlations in model responses induced by object motion. The responses of cortical neurons typically exhibit weakly correlated noise, such that neurons with similar tuning generally have positively correlated noise (i.e., "limited range" correlations, Cohen and Kohn 2011). This is true for heading-selective neurons in areas MSTd and VIP (Chen et al. 2013; Gu et al. 2011). In the simulations described thus far, we did not specifically introduce any correlated noise among model neurons: neural responses were generated with independent Poisson noise. Importantly, however, task-irrelevant variations in the stimulus induce correlations into the neural population response for a given task variable. For example, in estimating heading, what matters are the correlations among neurons for each particular heading, irrespective of the direction of object motion. Note that this notion of response correlation (conditioned only on the task-relevant stimulus) falls somewhere between what are commonly known as signal correlations (averaged over all stimuli) and noise correlations (conditioned on all stimulus variables). For performing marginalization, these response correlations induced by task-irrelevant variables may be an important constraint on performance.

To examine the correlation structure among responses of the model neurons, we computed the average heading-conditional covariance matrix of neural responses to the training set of stimuli (see METHODS). Figure 12*A* shows the correlation structure among responses of the model neurons that were used to train ALM. The *top left* and *bottom right* quadrants of Fig. 12*A* show response correlations for pairs of neurons with matched multisensory congruency: congruent-congruent pairs and opposite-opposite pairs. The *top right* and *bottom left* quadrants show correlations for pairs with mismatched congruency (congruent-opposite pairs). Within each quadrant of Fig. 12*A*, neurons are sorted according to their visual heading preferences. A clear diagonal pattern is seen in Fig. 12*A*, such that neurons with similar visual heading preferences tend to be positively correlated, and neurons with widely discrepant heading preferences tend to be negatively correlated. This pattern of correlations is similar to the "limited range" correlation structure typically seen among cortical neurons. Intuitively, this pattern arises because a moving object will have similar effects on the responses of two neurons with similar visual heading tuning, and the object will have different effects on responses of two neurons with discrepant visual heading preferences.

Figure 12*B* shows the correlation structure for the model neurons again, but with neurons sorted according to their vestibular heading preferences. The result is similar for congruent-congruent pairs and opposite-opposite pairs, with positive correlations along the diagonal. However, for pairs with mismatched congruency, object motion induces negative correlations for pairs with similar vestibular heading preferences and positive correlations for pairs with opposite vestibular preferences. This is expected because congruent-opposite pairs with opposite vestibular heading preferences would have matched visual heading preferences, and the effect of object motion on neural responses should be determined by the similarity of visual heading tuning.

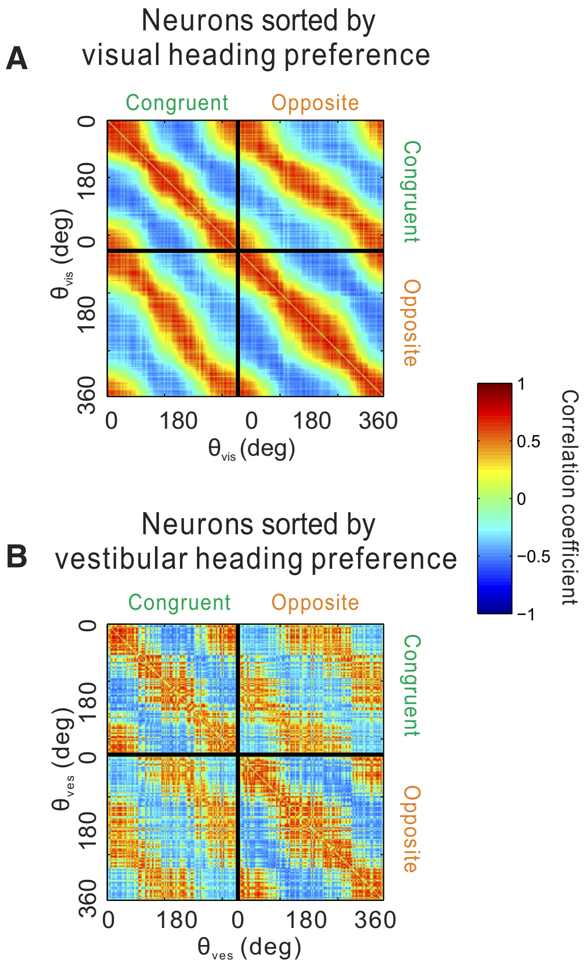

Fig. 12. Pairwise correlations between model neurons induced by moving objects (object probability = 1; see METHODS for details). *A*: average heading-conditional correlations between model neurons, sorted by visual heading preference. The correlation coefficient between each pair of neurons is color-coded. Neurons are grouped according to whether they have congruent or opposite heading preferences based on visual and vestibular inputs. The diagonal structure reveals that responses are positively correlated for pairs of model neurons with similar visual heading preferences. *B*: the same correlation data, with neurons sorted by vestibular heading preference. Note that congruent-opposite pairs of neurons (*top right* and *bottom left* quadrants) tend to have negative correlations when their vestibular heading preferences are similar.

Thus, while we did not intentionally introduce any correlated noise into our model neurons, population responses used to perform marginalization exhibit strong correlations that are induced by object motion. Given the strength of the correlations induced by object motion, we expected that introducing additional correlated noise into the model would have modest effects on the performance of ALM. To examine this possibility, we took advantage of the fact that previous studies have quantified the relationship between noise correlations and signal correlations for pairs of simultaneously recorded neurons from area MSTd (Gu et al. 2011, 2014). When we built this empirically observed pattern of correlated noise into responses of our model neurons, we found that it increased the RMS heading errors somewhat (black stars in Fig. 7, *A* and *B*), but did not fundamentally alter the pattern of decoding weights that were learned by ALM (data not shown). Our simulations suggest that task-irrelevant variables (such as object motion) may provide the dominant source of neural response correlations. Thus, when marginalization is a necessary operation, information in population responses may be substantially more constrained by the effects of task-irrelevant variables than by intrinsic noise correlations. This issue deserves further study using large populations of simultaneously recorded neurons.

DISCUSSION

We have addressed a fundamental question regarding how the brain can dissociate variables that are confounded in the sensory input. We focused on an example challenge of this form: identifying self-motion in the presence of object motion. We assumed a situation in which the visual responses of cortical neurons cannot distinguish self-motion and object motion, and we demonstrate that decoding a mixture of congruent and opposite cells according to their vestibular tuning allows a substantial reduction in heading errors when moving objects are present. We further show, through theory and simulations, that the strategy of decoding multisensory neurons by their vestibular preferences is similar to the decoding weights that are determined by an optimal linear approximation to marginalization (Fig. 7). Finally, we demonstrate that such a decoding strategy could be learned through simple reinforcement learning mechanisms, thus making it quite biologically plausible. Together, our findings reveal novel computational strategies by which the brain may dissociate self-motion and object motion by making use of multisensory neurons with variable congruency between visual and vestibular tuning. These strategies may have general applicability to other problems of marginalization and causal inference that the brain needs to solve.

It is important to note that we did not attempt to model the detailed interactions between background motion and object motion in the visual images, nor how these stimuli interact with the receptive field properties of model neurons. This approach was taken for three reasons. First, sufficient physiological information is not currently available to model in detail how neural responses depend on both object motion and self-motion. Second, by assuming that visual responses confound self-motion and object motion, we create a worst-case scenario in which marginalization is not possible based on visual responses alone. If a computational approach to marginalization succeeds in this case, it should also work well in the case of other (less severe) visual interactions between self-motion and object motion. Third, by simply modeling self-motion and object motion as vectors (rather than detailed patterns of image motion), our approach is very general and can be readily applied to other sensory systems and tasks. In addition, our approach would generalize to the case of multiple moving objects in a scene. If the visual responses of the model neurons are tuned to the vector average direction of self-motion and the motion of multiple objects, then our approach to estimating heading would work equally well for the case of multiple objects.

ALM. Marginalization is a common computation that the brain needs to perform (Kersten et al. 2004; Ma 2010). We often want our behavior to be governed by one or a few stimulus parameters while ignoring other “nuisance” parameters. In such circumstances, we may want to compute the probability distribution over a parameter of interest, while marginalizing out the nuisance parameters.

Recently, Beck et al. (2011) have shown that a neural network having quadratic nonlinearities with divisive normalization can perform exact marginalization over linear combinations of stimulus variables when the sensory neurons are directly tuned to both stimuli. Marginalization of more general stimulus variables and dependencies requires more sophisticated nonlinearities. Although it may be possible to implement such nonlinear computations exactly, the complexity of this nonlinear code would likely grow with the dimensionality and complexity of the joint posterior (e.g., number of moving objects). Thus we have chosen to focus on finding a simple, linear decoder that could achieve a close approximation to the marginalized posterior distribution. It was not clear a priori how well such a linear decoder would perform, but our results show that heading estimates obtained by ALM are quite robust to object motion.

The pattern of decoding weights obtained by ALM is often similar to the vestibular tuning of multisensory neurons. However, fine differences between the learned weights and the vestibular tuning curves are clearly important for achieving a good approximation to the marginal posterior when the population has diverse tuning properties (Figs. 5–7). It is possible that the vestibular tuning of MSTd neurons is used as an initial estimate of the decoding weights, which are then refined through learning in the presence of object motion.

Although we have applied ALM to the specific problem of dissociating self-motion and object motion, it is important to note that our formulation of ALM is very general, and it should be useful for solving a variety of problems that the brain faces. Importantly, this approach to marginalization may be particularly important in cases (such as object motion) for which there are no internal signals (e.g., efference copy of motor commands) that can be used to dissociate sensory input into components related to different physical causes.

Visual and vestibular contributions to dissociating self-motion and object motion. Our model assumes that the visual responses of neurons inherently confound velocity vectors associated with self-motion and object motion, such that visual heading perception is biased in the presence of moving objects. Of course, if the visual system is capable of partially discounting object motion on its own, without vestibular input, then this reduces the burden that is placed on a mechanism such as we propose. Psychophysical studies have shown that humans' visual perception of heading can be biased by object motion, especially when the object crosses the focus of expansion in a flow field (Dokka et al. 2015a; Fajen and Kim 2002; Layton and Fajen 2016; Royden and Hildreth 1996; Warren and Saunders 1995). This suggests that visual processing only partially discounts object motion. A variety of computational models have been proposed in which neural processing of visual motion signals may account for the observed biases in heading perception induced by moving objects (e.g., Hanada 2005; Layton et al. 2012; Royden 2002). It is important to note that our proposed multisensory mechanism is complementary to the existence of purely visual mechanisms that may also contribute to discounting object motion during self-motion perception.

Indeed, the extent to which visual mechanisms alone can dissociate self-motion and object motion is currently somewhat unclear. Recent behavioral studies suggest that the visual system has mechanisms for parsing retinal image motion into components related to self-motion and object motion (Matsumiya and Ando 2009; Warren and Rushton 2007, 2009a, 2009b), and some species appear to possess mechanisms for isolating local motion from background motion as early as the retina (Olveczky et al. 2003), but they do not establish that such mechanisms can fully tease apart motion components resulting from observer and object motion. It seems likely that nonvisual cues that accompany self-motion, such as signals arising from the otolith organs of the vestibular system (Angelaki and Cullen 2008), would provide valuable inputs for solving this computational problem. Indeed, a recent study of heading perception in macaque monkeys has demonstrated that vestibular signals dramatically reduce biases in heading percepts caused by object motion (Dokka et al. 2015a), consistent with the predictions of our model. In addition, recent psychophysical studies have explored the complementary question of how nonvisual signals related to self-motion (including vestibular signals) contribute to perception of object motion (Dokka et al. 2015b; Fajen and Matthis 2013; MacNeilage et al. 2012), and these studies demonstrate that nonvisual inputs play important roles in discounting self-motion during perception of object motion. In this respect, it is worth noting that our ALM approach could be used just as well to marginalize over heading and estimate object motion.

Plausibility and generality of proposed decoding schemes. Is it plausible that responses of both congruent and opposite neurons are decoded according to their vestibular tuning? A previous study (Gu et al. 2008) examined choice-related activity of MSTd neurons during a heading discrimination task by measuring choice probabilities (Britten et al. 1996; Nienborg et al. 2012), which reflect both the readout weights applied to neural activity as well as correlated noise among neurons (Haefner et al. 2013). When heading discrimination was based on vestibular cues, both congruent and opposite cells responded more strongly when the monkey reported a heading consistent with the neurons' vestibular heading preference. However, when heading perception was based solely on optic flow, opposite cells responded less strongly when the monkey reported in favor of their visual heading preference, thus yielding choice probabilities that were significantly <0.5 (Fig. 6c of Gu et al. 2008). As suggested by Gu et al. (2008) and recently confirmed through simulations (Gu et al. 2014), this finding is consistent with the idea that responses of opposite cells are decoded according to their vestibular tuning. Thus the data of Gu et al. (2008) may reflect the same general decoding strategy that affords robustness to object motion.

Is a strategy that involves selective decoding of congruent and opposite cells specific to visual-vestibular integration? Other recent results suggest that this may be a more general phenomenon. In area MT, many neurons show selectivity for depth based on binocular disparity (DeAngelis and Uka 2003; Maunsell and Van Essen 1983) or motion parallax (Nadler et al. 2008) cues. Interestingly, almost one-half of these neurons have incongruent depth tuning for disparity and motion parallax (Kim et al. 2015; Nadler et al. 2013). Our laboratory suggested recently (Nadler et al. 2013) that these incongruent neurons may play an important role in detecting object motion in the world during self-motion, and preliminary results support this hypothesis (Kim et al. 2014). Thus neural representations involving cells with incongruent tuning for different stimulus parameters may provide valuable inputs to computations involving marginalization or causal inference.

Testable predictions. This study makes some important predictions that can be tested experimentally. First, it predicts that linearly decoding MSTd neurons according to their vestibular tuning, as implemented via the ALM decoder, should produce heading estimates that are more robust to object motion than those obtained from other linear decoding strategies. This can be tested by measuring responses of MSTd neurons to many combinations of directions of self-motion and object motion and then decoding responses according to different strategies. Preliminary results from our laboratory suggest that this is indeed the case (Sasaki et al. 2013).

Second, our results suggest some tradeoff between marginalization and the potential sensitivity gains available through cue integration. Training ALM with moving objects reduces heading biases caused by object motion, but also modestly reduces the gains in heading sensitivity that can be achieved via cue integration (Fig. 7*D*). This suggests that the benefit of integrating optic flow and vestibular cues for heading discrimination (Gu et al. 2008) may be somewhat reduced when heading discrimination tasks are performed in the presence of moving objects, if the brain relies on linear responses transformations such as ALM. Recent results, although inconclusive, are broadly consistent with this prediction (Dokka et al. 2015a).

Third, our simulations suggest that response correlations induced by task-irrelevant variables (such as object motion) may play a major role in limiting the information conveyed by a neural population, perhaps substantially more so than the commonly observed limited-range noise correlations. It will be useful to directly compare the strengths of these different sources of response correlations in empirical data sets involving large neural populations.

In closing, we have provided a biologically plausible mechanism by which visual and vestibular signals can be combined to dissociate self-motion and object motion information. Our approach to performing ALM may be valuable in a variety of neural computations implemented in both sensory and motor systems.

ACKNOWLEDGMENTS

We thank Jeff Beck for helpful conversations.

Present address of H. R. Kim: Dept. of Molecular and Cell Biology, Harvard University, Cambridge, MA 02138.

GRANTS

The work was supported by National Eye Institute (NEI) Grants EY-016178 and EY-013644, by an NEI CORE Grant (EY-001319), and by a T32 fellowship (T32DC009974) and grant from the McNair Foundation to X. Pitkow.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

H.R.K., X.P., D.E.A., and G.C.D. conception and design of research; H.R.K. and X.P. analyzed data; H.R.K., X.P., D.E.A., and G.C.D. interpreted results of experiments; H.R.K. prepared figures; H.R.K. and G.C.D. drafted manuscript; H.R.K., X.P., D.E.A., and G.C.D. edited and revised manuscript; H.R.K., X.P., D.E.A., and G.C.D. approved final version of manuscript.

REFERENCES

Angelaki DE, Cullen KE. Vestibular system: the many facets of a multimodal sense. *Annu Rev Neurosci* 31: 125–150, 2008.

Beck JM, Latham PE, Pouget A. Marginalization in neural circuits with divisive normalization. *J Neurosci* 31: 15310–15319, 2011.

Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

Bremmer F, Klam F, Duhamel JR, Ben Hamed S, Graf W. Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16: 1569–1586, 2002.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* 13: 87–100, 1996.

Chen A, DeAngelis GC, Angelaki DE. A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. *J Neurosci* 31: 3082–3094, 2011a.

Chen A, DeAngelis GC, Angelaki DE. Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *J Neurosci* 31: 11617–11627, 2011b.

Chen A, DeAngelis GC, Angelaki DE. Functional specializations of the ventral intraparietal area for multisensory heading discrimination. *J Neurosci* 33: 3567–3581, 2013.

Chen A, DeAngelis GC, Angelaki DE. Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. *J Neurosci* 31: 12036–12052, 2011c.

Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011.

Crowell JA, Banks MS, Shenoy KV, Andersen RA. Visual self-motion perception during head turns. *Nat Neurosci* 1: 732–737, 1998.

Dayan P, Abbott LF. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: Massachusetts Institute of Technology Press, 2001, p. xv, 460.

DeAngelis GC, Uka T. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol* 89: 1094–1111, 2003.

Dokka K, DeAngelis GC, Angelaki DE. Multisensory integration of visual and vestibular signals improves heading discrimination in the presence of a moving object. *J Neurosci* 35: 13599–13607, 2015a.

Dokka K, Macneilage PR, DeAngelis GC, Angelaki DE. Multisensory self-motion compensation during object trajectory judgments. *Cereb Cortex* 25: 619–630, 2015b.

Duffy CJ. MST neurons respond to optic flow and translational movement. *J Neurophysiol* 80: 1816–1827, 1998.

Ecker AS, Berens P, Tolias AS, Bethge M. The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci* 31: 14272–14283, 2011.

Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–433, 2002.

Fajen BR, Kim NG. Perceiving curvilinear heading in the presence of moving objects. *J Exp Psychol Hum Percept Perform* 28: 1100–1119, 2002.

Fajen BR, Matthis JS. Visual and non-visual contributions to the perception of object motion during self-motion. *PLoS One* 8: e55446, 2013.

Fajen BR, Parade MS, Matthis JS. Humans perceive object motion in world coordinates during obstacle avoidance. *J Vis* 13: 25, 2013.

Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* 15: 146–154, 2012.

Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci* 29: 15601–15612, 2009.

Gibson JJ. The perception of visual surfaces. *Am J Psychol* 63: 367–384, 1950.

Gu Y, Angelaki DE, DeAngelis GC. Contribution of correlated noise and selective decoding to choice probability measurements in extrastriate visual cortex. *Elife* 3: e02670, 2014.

Gu Y, Angelaki DE, DeAngelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* 11: 1201–1210, 2008.

Gu Y, DeAngelis GC, Angelaki DE. A functional link between area MSTd and heading perception based on vestibular signals. *Nat Neurosci* 10: 1038–1047, 2007.

Gu Y, Fetsch CR, Adeyemo B, DeAngelis GC, Angelaki DE. Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* 66: 596–609, 2010.

Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, Sunkara A, DeAngelis GC, Angelaki DE. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* 71: 750–761, 2011.

Gu Y, Watkins PV, Angelaki DE, DeAngelis GC. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J Neurosci* 26: 73–85, 2006.

Haefner RM, Gerwinn S, Macke JH, Bethge M. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat Neurosci* 16: 235–242, 2013.

Hanada M. Computational analyses for illusory transformations in the optic flow field and heading perception in the presence of moving objects. *Vision Res* 45: 749–758, 2005.

Hinton GE, Dayan P. Varieties of Helmholtz machine. *Neural Netw* 9: 1385–1403, 1996.

Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. *Nat Neurosci* 9: 690–696, 2006.

Kersten D, Mamassian P, Yuille A. Object perception as Bayesian inference. *Annu Rev Psychol* 55: 271–304, 2004.

Kim HR, Angelaki DE, DeAngelis GC. Detecting moving objects based on cue conflict between disparity and motion parallax: behavior and physiology. *Soc Neurosci Abstr* 726: 717, 2014.

Kim HR, Angelaki DE, DeAngelis GC. A novel role for visual perspective cues in the neural computation of depth. *Nat Neurosci* 18: 129–137, 2015.

Law CT, Gold JI. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat Neurosci* 12: 655–663, 2009.

Layton OW, Fajen BR. The Temporal Dynamics of Heading Perception in the Presence of Moving Objects. *J Neurophysiol* 115: 286–300, 2016.

Layton OW, Mingolla E, Browning NA. A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *J Vis* 12: 20, 2012.

Logan DJ, Duffy CJ. Cortical area MSTd combines visual cues to represent 3-D self-movement. *Cereb Cortex* 16: 1494–1507, 2006.

Ma WJ. Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Res* 50: 2308–2319, 2010.

Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. *Nat Neurosci* 9: 1432–1438, 2006.

MacNeilage PR, Zhang Z, DeAngelis GC, Angelaki DE. Vestibular facilitation of optic flow parsing. *PLoS One* 7: e40264, 2012.

Matsumiya K, Ando H. World-centered perception of 3D object motion during visually guided self-motion. *J Vis* 9: 15, 2009.

Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49: 1148–1167, 1983.

Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A. Information-limiting correlations. *Nat Neurosci* 17: 1410–1417, 2014.

Morgan ML, DeAngelis GC, Angelaki DE. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59: 662–673, 2008.

Nabney IT. *Netlab: Algorithms for Pattern Recognition*. London: Springer, 2002.

Nadler JW, Angelaki DE, DeAngelis GC. A neural representation of depth from motion parallax in macaque visual cortex. *Nature* 452: 642–645, 2008.

Nadler JW, Barbash D, Kim HR, Shimpi S, Angelaki DE, DeAngelis GC. Joint representation of depth from motion parallax and binocular disparity cues in macaque area MT. *J Neurosci* 33: 14061–14074, 2013.

Nienborg H, Cohen MR, Cumming BG. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu Rev Neurosci* 35: 463–483, 2012.

Olveczky BP, Baccus SA, Meister M. Segregation of object and background motion in the retina. *Nature* 423: 401–408, 2003.

Pearl J. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2000.

Royden CS. Computing heading in the presence of moving objects: a model that uses motion-opponent operators. *Vision Res* 42: 3043–3058, 2002.

Royden CS, Banks MS, Crowell JA. The perception of heading during eye movements. *Nature* 360: 583–585, 1992.

Royden CS, Hildreth EC. Human heading judgments in the presence of moving objects. *Percept Psychophys* 58: 836–856, 1996.

Rushton SK, Bradshaw MF, Warren PA. The pop out of scene-relative object movement against retinal motion due to self-movement. *Cognition* 105: 237–245, 2007.

Sasaki R, Angelaki DE, DeAngelis GC. Estimating heading in the presence of moving objects: Population decoding of activity from area MSTd. *Soc Neurosci Abstr* 360: 317, 2013.

Schlack A, Hoffmann KP, Bremmer F. Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16: 1877–1886, 2002.

Shamir M, Sompolinsky H. Implications of neuronal diversity on population coding. *Neural Comput* 18: 1951–1986, 2006.

Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998, p. xviii, 322.

Warren PA, Rushton SK. Optic flow processing for the assessment of object movement during ego movement. *Curr Biol* 19: 1555–1560, 2009a.

Warren PA, Rushton SK. Perception of object trajectory: parsing retinal motion into self and object movement components. *J Vis* 7: 2, 2007.

Warren PA, Rushton SK. Perception of scene-relative object movement: optic flow parsing and the contribution of monocular depth cues. *Vision Res* 49: 1406–1419, 2009b.

Warren WH Jr, Saunders JA. Perceiving heading in the presence of moving objects. *Perception* 24: 315–331, 1995.

Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143, 1994.