J Comput Neurosci
DOI 10.1007/s10827-014-0545-1

Neural representation of probabilities for Bayesian inference

Dylan Rich · Fanny Cazettes · Yunyan Wang · José Luis Peña · Brian J. Fischer

Received: 3 April 2014 / Revised: 7 December 2014 / Accepted: 23 December 2014

Abstract Bayesian models are often successful in describing perception and behavior, but the neural representation of probabilities remains in question. There are several distinct proposals for the neural representation of probabilities, but they have not been directly compared in an example system. Here we consider three models: a non-uniform population code where the stimulus-driven activity and distribution of preferred stimuli in the population represent a likelihood function and a prior, respectively; the sampling hypothesis which proposes that the stimulus-driven activity over time represents a posterior probability and that the spontaneous activity represents a prior; and the class of models which propose that a population of neurons represents a posterior probability in a distributed code. It has been shown that the non-uniform population code model matches the representation of auditory space generated in the owl's external nucleus of the inferior colliculus (ICx). However, the alternative models have not been tested, nor have the three models been directly compared in any system. Here we tested the three models in the owl's ICx. We found that spontaneous firing rate and the average stimulus-driven response of these neurons were not consistent with predictions of the sampling hypothesis. We also found that neural activity in ICx under varying levels of sensory noise did not reflect a posterior probability. On the other hand, the responses of ICx neurons were consistent with the non-uniform population code model. We further show that Bayesian inference can be implemented in the non-uniform population code model using one spike per neuron when the population is large and is thus able to support the rapid inference that is necessary for sound localization.

Keywords Bayesian inference · Neural coding · Sound localization · Barn owl · Population code

Action Editor: B. A. Olshausen

D. Rich · B. J. Fischer (✉)
Department of Mathematics, Seattle University, 901 12th Ave, Seattle, WA 98122, USA
e-mail: fischer9@seattleu.edu

F. Cazettes · Y. Wang · J. L. Peña
Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA

1 Introduction

Perception and behavior are often consistent with Bayesian inference. Bayesian models describe how sensory and prior information can be combined optimally (Fig. 1). Bayesian inference about a variable θ, such as the direction of a sound source, based on sensory information S is performed using the posterior probability $p(\theta|S)$. The posterior probability captures what is known about the variable θ after sensory input S has been observed. Bayes' rule tells us that the posterior probability can be written as $p(\theta|S) \propto p(S|\theta)p(\theta)$, where $p(\theta)$ is the prior and the probability $p(S|\theta)$ viewed as a function of θ is called the likelihood. The likelihood describes how the sensory input depends on the external variable θ. The prior describes the probability distribution of θ before any sensory information has been received. Bayesian models highlight the importance of prior distributions in solving perceptual problems that rely on ambiguous sensory information (Fischer and Peña 2011; Weiss et al. 2002). For example, visual information from small regions of an image does not uniquely signal the velocity of an object. A Bayesian model with a prior that favors slow velocities explains human perception of visual motion, including misjudgments of velocities, based on the ambiguous local motion information (Weiss et al. 2002). An important open question is, however, how the components of a Bayesian model are represented in the brain.

Published online: 06 January 2015

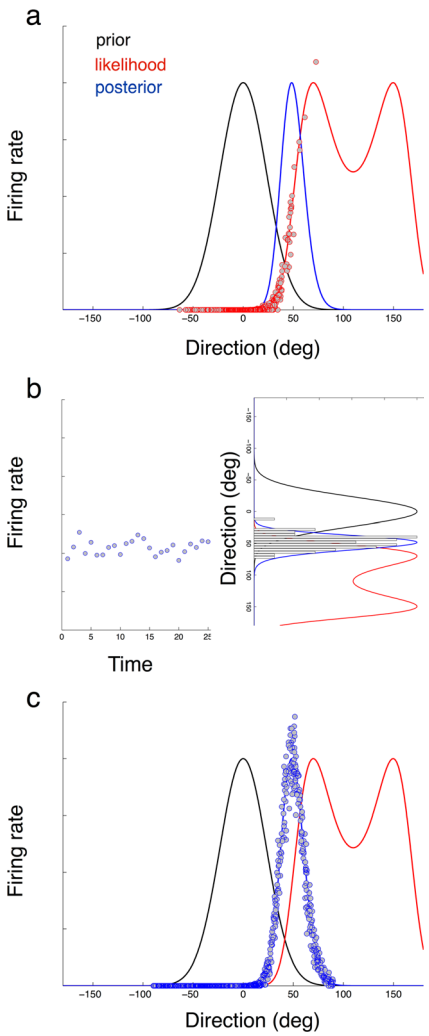

Fig. 1 Models for the neural implementation of Bayesian inference **a** The non-uniform population code model of Fischer and Peña (2011) proposes that the neural activity encodes the likelihood (*red*) and the prior is represented in the distribution of preferred directions of midbrain neurons. **b** The sampling hypothesis predicts that neural activity represents samples from the posterior. A histogram of the neural activity over time approximates the posterior. **c** A population may directly encode the posterior in the activities of neurons that are uniformly distributed

Several models exist for the neural representation of a probability distribution. Here we consider models that make specific predictions about how neural circuits represent a probability defined on environmental variables θ and the resulting sensory input S. These models differ in the way that the response of a neuron or group of neurons encodes a posterior probability, and thus make experimentally testable predictions that allow the models to be evaluated and distinguished between.

The non-uniform population code model proposes that the likelihood is represented in the shape of the tuning curves and the prior is represented in the distribution of preferred stimuli (Fig. 1a) (Shi and Griffiths 2009; Fischer and Peña 2011; Girshick et al. 2011). The representation of the prior $p(\theta)$ in the distribution of preferred stimuli means that more neurons will have preferred stimuli θ_n in regions of the stimulus space that have high prior probability. This is achieved in the model by having the preferred stimuli sampled from the prior. The model prediction that neural responses are determined by the likelihood function means that the tuning of a neuron with preferred stimulus θ_n to the sensory input S is determined by the statistics of the sensory input $p(S|\theta_n)$. Specifically, under this model the activity of neurons in the population to stimulus S is assumed to be proportional to the likelihood $p(S|\theta)$. This model is consistent with experimental observations in the owl's midbrain (Fischer and Peña 2011).

The sampling hypothesis for the neural implementation of Bayesian inference states that neurons correspond to variables, and the stochastic firing encodes the distribution of the variables (Fig. 1b). Given a stimulus S, a neuron corresponding to variable θ will have firing rates that are samples from the posterior probability $p(\theta|S)$ (Berkes et al. 2011; Fiser et al. 2010). Thus, a neuron encoding the posterior probability $p(\theta|S)$ will have a high mean firing rate when a large value of the variable θ is most probable and will have a low mean firing rate when a small value of θ is most probable. Also, the firing rate will have high variability over time when the posterior has a large variance and will have low variability over time when the posterior has a small variance. When the firing rate reflects the posterior in this way, a histogram of stimulus-driven activity over time will approximate the posterior $p(\theta|S)$ (Fig. 1b). If there is no stimulus, the probability over θ is the prior distribution $p(\theta)$. Thus, the sampling hypothesis predicts that spontaneous activity, in the absence of a stimulus, is sampled from the prior (Berkes et al. 2011; Fiser et al. 2010). This implies that the spontaneous activity will be clustered around the values of θ that have high prior probability, with a variability that reflects the prior variance. The sampling hypothesis therefore leads to the prediction that the distribution of spontaneous activity of a neuron is the average of the distribution of the stimulus driven activity. These predictions are consistent with observations in ferret V1 (Berkes et al. 2011).

A third hypothesis is that neural activity in a population directly encodes the posterior or the log of the posterior (Fig. 1c) (Anderson and Van Essen 1994; Barber et al. 2003; Eliasmith and Anderson 2004; 1993; Gold and Shadlen 2000; Sahani and Dayan 2003; Sanger 1996; Simoncelli 2003, 2009). In models of this type, neurons have preferred stimuli θ_n that uniformly cover the range of possible stimuli, regardless of the prior distribution. The pattern of stimulus-driven activity of neurons in the population encodes the posterior $p(\theta|S)$ such that the firing rate of a neuron with preferred stimulus θ_n reflects the value of the posterior at or near θ_n. In this way, the pattern of activity across the population matches the shape of the posterior distribution. Specifically,

neurons with preferred stimuli θ_n that have high posterior probability will have high firing rates, and neurons with preferred stimuli θ_n that have low posterior probability will have low firing rates.

These models for the neural implementation of Bayesian inference have not been directly compared in an example system. The rapid sound localization behavior of the barn owl for brief sounds is consistent with a Bayesian model (Fischer and Peña 2011). Here, the sensory information is the interaural time difference (ITD), which is an ambiguous cue for direction (Fischer and Peña 2011). A prior that emphasizes central directions encodes the owl's behavioral bias and allows for accurate localization for ethologically relevant directions. The neural implementation of Bayesian inference is consistent with a non-uniform population code model where the prior is represented in the distribution of preferred stimuli. The predictions of the non-uniform population code model match the representation of auditory space generated in the owl's external nucleus of the inferior colliculus (ICx) (Fischer and Peña 2011). However, the alternative models have not been tested in the owl, nor have the three models been directly compared in any system. Here we test alternative models for the neural implementation of Bayesian inference in the owl's midbrain. Further, we test whether any of the models that are consistent with the data for the owl are able to implement Bayesian inference with few spikes per neuron, as required for the fast temporal dynamics of the owl's sound localization.

2 Materials and methods

2.1 Neurophysiology

Data collection methods have been described in detail previously (Peña and Konishi 2002; Wang et al. 2012). Procedures complied with the U.S. National Institutes of Health and the Albert Einstein College of Medicine's Institute of Animal Studies guidelines. Briefly, adult barn owls (*Tyto alba*) of both sexes were restrained with a soft cloth jacket and anesthetized with intramuscular injections of ketamine hydrochloride (20 mg/kg Ketaset) and xylazine (4 mg/kg Anased).

Single units were extracellularly recorded in ICx using 1 MΩ tungsten electrodes (A-M Systems). Single units were isolated by spike amplitude. Spikes were detected offline using spike-discrimination software written in MATLAB (MathWorks). ICx was located stereotaxically (Knudsen 1983) and by the unambiguous response of its neurons to interaural time (ITD) and level (ILD) differences.

All free-field experiments were performed in a double-walled sound-attenuating chamber (Industrial Acoustics) lined with echo-absorbing acoustical foam (Sonex). The free-field spatial tuning of the neurons was measured using a custom-made hemispherical array of 144 speakers (Sennheiser, 3P127A) constructed inside the sound-attenuating chamber. The speaker array range is ±100° in azimuth and ±80° in elevation. The angular separation between the speakers varied from 10° to 30°. During recordings, the owl was placed in a stereotaxic frame at the center of the speaker array such that all speakers were equidistant (60 cm) from the owl's head. Auditory stimuli used to measure free-field spatial RFs consisted of five repetitions of 100 ms duration broadband signals (0.5–10 kHz), with a 5 ms rise–fall time and 50 dB intensity. Stimuli were presented with an interstimulus interval (ISI) of 300 ms in which speaker location was randomized.

Extracellular ITD tuning was measure at different levels of binaural correlation (BC). Acoustic stimuli were digitally synthesized with a personal computer and delivered to both ears by calibrated earphone assemblies. Auditory stimuli consisted of broadband noise bursts [0.5–12.0 kHz; 50 or 100 msec in duration and 5 msec rise and decay times; sound level of 40–50 dB sound pressure level]. The computer synthesized three random noises to obtain different values of binaural correlation. One of them was delivered to one ear and its copy to the other ear, making the correlated component of the stimulus. Each of the other two noises was used as the uncorrelated component of the stimulus by adding it to the correlated noise while keeping the sound level constant. Binaural correlation varies with the relative amplitude of the uncorrelated and correlated noises by $1/(1 + k^2)$, where k is the ratio between the root-mean-square amplitudes of the uncorrelated and correlated noises.

Intracellular responses were recorded *in vivo* using sharp borosilicate glass electrodes filled with 2 M potassium acetate and 4 % neurobiotin. Analog signals were amplified (Axoclamp 2A; Axon Instruments, Foster City, CA) and stored in the computer. We identified ICx neurons by labeling their axons, which project to the optic tectum. The tracer neurobiotin was injected by iontophoresis (3 nA positive, 300 msec current steps; three per second for 5–30 min). After the experiment, the owls were overdosed with Nembutal and perfused with 2 % paraformaldehyde. Brain tissue was cut in 60-μm-thick sections and processed according to standard protocols (Kita and Armstrong 1991). Membrane potentials were median-filtered using a sliding window of 1 ms to remove spikes for the analysis of variability over time.

2.2 Model

Our tests of the three models for the neural implementation of Bayesian inference utilize the Bayesian behavioral model of Fischer and Peña (2011). In this model, the environmental variable is the azimuth of a sound source and the sensory input is the ITD. The prior in the Bayesian model of the owl's localization behavior is a Gaussian with zero mean and a standard deviation of 23.3 deg. This is consistent with the

owl's localization performance and also with the experimental observation that the owl spends most of the time with the prey in front of it while engaged in prey capture (Edut and Eilam 2004). The prior in the description of the localization of brief sounds corresponds to the owl's behavioral bias. While the distribution of initial prey directions may be uniform, the Gaussian prior represents the ethological importance of accurate localization when the prey is in a position for the owl to strike (Fischer and Peña 2011; Salinas 2011). The likelihood is based on the model that ITD is a sinusoidal function of direction, corrupted by Gaussian noise with zero mean and standard deviation 41.2 μs. This model is constrained by direct measurements of the ITD in the signals in the ear canals for sounds arising from different directions in space (Hausmann et al. 2009). This model describes the owl's rapid localization of brief sounds and this is the context in which our tests are made.

We modified the neural model of Fischer and Peña (2011) to consider the case of computation with single spikes per neuron. The tuning curves of the model neurons are given by

$$a_n(\theta) = a_{max}\exp\left(-\frac{(Asin(\omega\theta)-Asin(\omega\theta_n))^2}{2\sigma(BC)^2}\right)$$

where a_{max} is the maximum firing rate, θ_n is the preferred direction, $\sigma(BC)$ determines the width of the curve depending on the binaural correlation BC, and $A=260\ \mu s$ and $\omega=0.143\ rad$ are determined by the head-related transfer function (Hausmann et al. 2009). The width parameter of the tuning curve is given by $\sigma(BC) = 219.38\exp(-11.31BC) + 41.2\,\mu s$ where the binaural correlation BC ranges from 0 for uncorrelated sounds to 1 for correlated sounds (Fischer and Peña 2011) The maximum firing rate a_{max} was selected to produce less than one spike per neuron on average. Spike counts were independent Poisson variables with the rates $a_n(\theta)$. Although ICx neurons likely have correlated spike count variability, the population vector decoder we use is robust to the presence of correlated variability (Fischer and Peña 2011) and the assumption of independence greatly reduces the computational complexity.

3 Results

3.1 Test of the sampling hypothesis in ICx

The sampling hypothesis for the neural implementation of Bayesian inference states that neurons correspond to variables θ and stochastic stimulus-driven firing encodes the posterior distribution of the variables $p(\theta|S)$, where S is the sensory input (Fig. 1b). Here, the environmental variable θ is the direction of a sound source and the sensory input S is the localization cue ITD. Similarly, the sampling hypothesis predicts that spontaneous activity is sampled from the prior $p(\theta)$ (Berkes et al. 2011; Fiser et al. 2010). This hypothesis leads to the prediction that the distribution of spontaneous activity of a neuron is the average of the distribution of the stimulus driven activity:

$$p(\theta) = \int p(\theta|S)p(S)dS.$$

This prediction for the distributions of spontaneous and stimulus-driven activity implies the simpler condition that the mean spontaneous firing rate is equal to the average over all stimuli of the trial-averaged stimulus-driven firing rate (Berkes et al. 2011; Fiser et al. 2010):

$$E[\theta] = \int E[\theta|S]p(S)dS.$$

Therefore, if the mean spontaneous firing rate is not equal to the mean stimulus-driven firing rate, the sampling hypothesis is not supported.

To test the sampling hypothesis in the owl, we compared the mean spontaneous firing rate of midbrain auditory neurons to the mean firing rate evoked by noise bursts presented in free field from different directions in space (Fig. 2). We computed the mean evoked activity by averaging over directions either using the Gaussian prior from the Bayesian model or a uniform distribution over directions. The uniform distribution matches the distribution of stimulus directions used in the experimental data collection. We found that, in contrast to the sampling hypothesis prediction (Berkes et al. 2011; Fiser et al. 2010), the mean spontaneous firing rate was significantly lower than the mean stimulus-driven firing rate, averaged over all directions (Kolmogorov-Smirnov $p<0.01$ for each neuron with both priors; $n=61$; Fig. 2). With spontaneous firing largely absent from midbrain neurons, the average of the stimulus driven activity cannot match the spontaneous activity, as required by the sampling hypothesis.

The sampling hypothesis also predicts that variability in neural responses over time represents uncertainty (Berkes et al. 2011; Fiser et al. 2010). Thus in the sampling hypothesis the neural variability over time should increase when sensory noise increases. In the owl's sound localization system, sensory noise increases when sounds at the two ears become decorrelated (Saberi et al. 1998). When independent noise is added to the ears, the correlation goes down, and it is said that the sounds are 'decorrelated'. The sensory noise increases by adding relatively greater amplitude noise signals to the left and right ears. We measured the variability of membrane potential

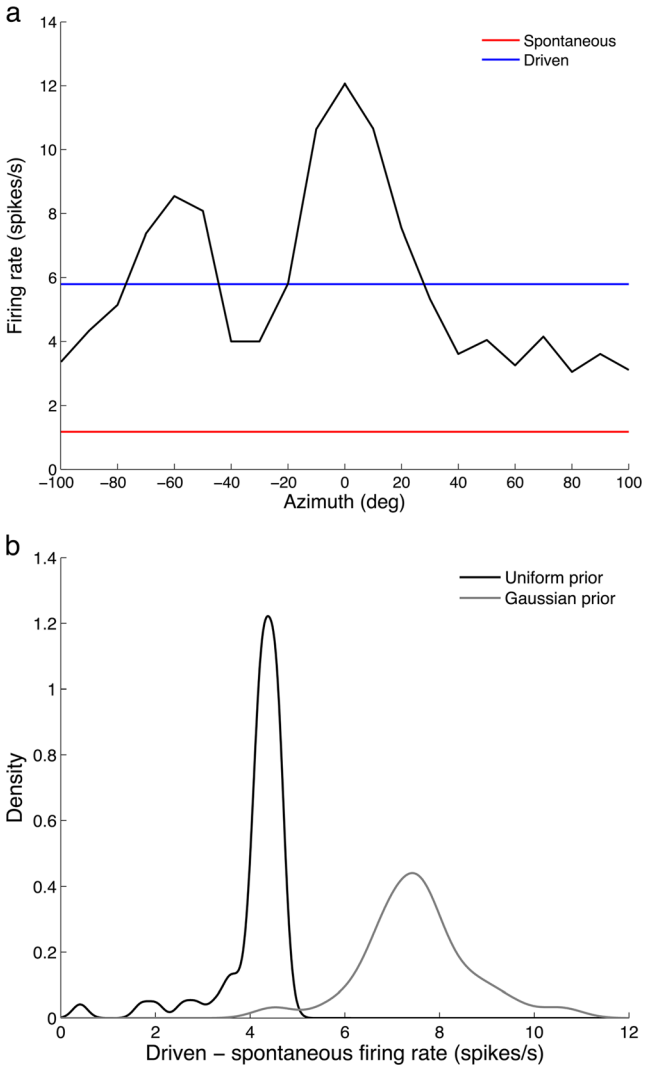

Fig. 2 Spontaneous and stimulus-driven activity in the midbrain (**a**) Auditory midbrain neurons in the owl have spatially restricted receptive fields (*black*) and low spontaneous activity. The spontaneous activity (*red*) is lower than the mean stimulus-driven activity (*blue*). (**b**) The mean spontaneous firing rate is significantly lower than the mean stimulus-driven firing rate, in contrast to the prediction of the sampling hypothesis. The mean stimulus-driven firing rate was computed using the Gaussian prior from the Bayesian model (*gray*) and the uniform distribution used in data collection (*black*)

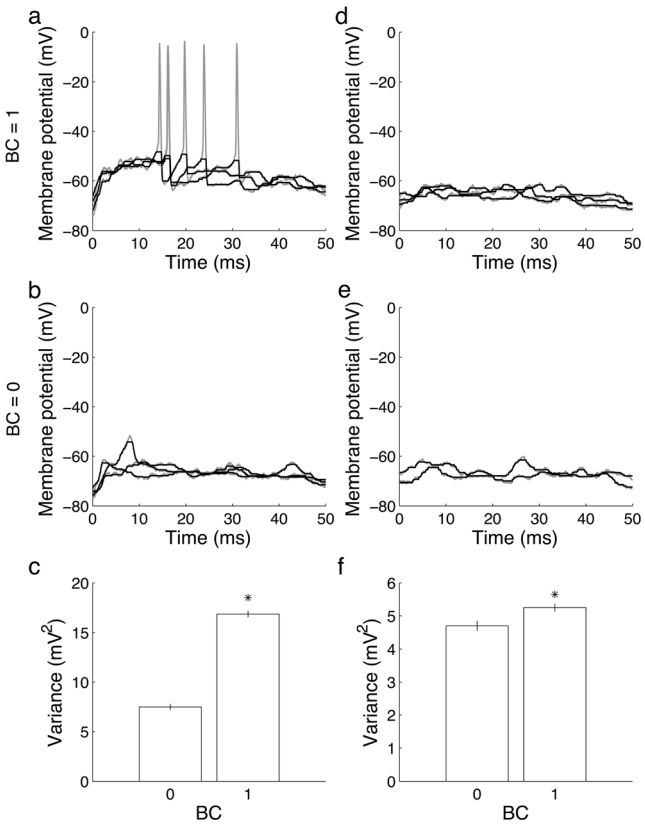

Fig. 3 Variability over time in midbrain responses (**a**–**f**) The variability over time in median-filtered (*black*) membrane potential responses is higher for correlated sounds (**a**,**d**) than for uncorrelated sounds (**b**,**e**), in contrast to the prediction of the sampling hypothesis. This occurs for stimulus conditions where spiking occurs (**a**–**c**) and where spiking is absent (**d**–**f**). *Error bars* in (**c**) and (**f**) are bootstrap standard deviations

responses of ICx neurons over time for correlated and uncorrelated sounds. We found that variability in the neural response at the preferred ITD and ILD was higher for correlated sounds than for uncorrelated sounds in 9 of 12 neurons (Fig. 3; $p<0.05$; *t*-test). This result is opposite to the prediction of the sampling hypothesis. Similarly, variability in the spike count at the preferred ITD and ILD was higher for perfectly correlated sounds than for sounds with zero correlation ($p=0.007$; *t*-test; $n=21$). The increase in variability of spike count responses we observed as stimulus correlation increases is consistent with Poisson-like neural noise and is expected due to the increase in mean firing rate as correlation increases. Consistent with our measurements in ICx, Saberi et al. (1998) reported a small but significant increase in variability in spiking responses of optic tectum neurons when correlation increases. Thus, the responses of midbrain auditory neurons in the owl do not support the sampling hypothesis.

3.2 Population coding of the posterior

Another hypothesis for the neural implementation of Bayesian inference is that firing rates of a population of neurons encode the posterior distribution (Fig. 1c). In these models, neurons have preferred stimuli θ_n covering the range of possible values. The firing rate of a neuron with preferred stimulus θ_n reflects the value of the posterior at or near θ_n (Anderson and Van Essen 1994; Barber et al. 2003; Eliasmith and Anderson 2004; 1993; Gold and Shadlen 2000; Sahani and Dayan 2003; Sanger 1996; Simoncelli 2003, 2009). We tested this hypothesis in the owl by comparing midbrain auditory responses with those predicted by the posterior distribution as binaural correlation changes.

The posterior probability is found by multiplying the likelihood function and the prior probability, and thus combines

sensory input with prior information. The relative weighting of sensory input and prior assumptions depends on the reliability of sensory input. As sensory input becomes unreliable, the posterior becomes more greatly influenced by the prior. The reliability of ITD as a sound localization cue is degraded when the binaural correlation decreases (Saberi et al. 1998). Consequently, in a Bayesian model, the posterior probability will shift towards the prior as binaural correlation decreases (Fig. 4a). The prediction for the representation of the posterior under changing binaural correlation is that the locus of maximal activity in the population would shift towards zero degrees as binaural correlation decreases. This prediction holds whether we assume that the firing rate of a neuron encodes the posterior directly or the log-posterior. In addition, this prediction of a shift in the locus of maximal activity in the population as BC decreases would hold for any model where neural activity is a fixed linear transformation of the posterior or log-posterior.

A neural representation of a shifting posterior as binaural correlation decreases can be accomplished by two possible

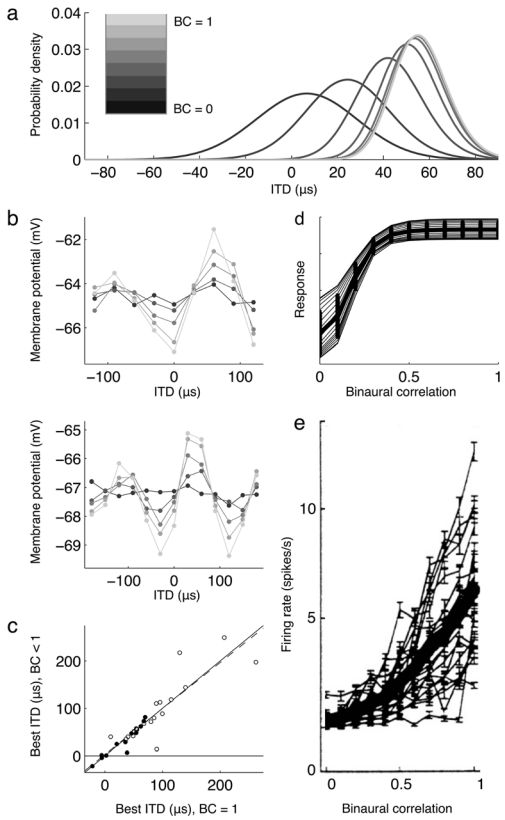

Fig. 4 Testing the representation of the posterior **a** The Bayesian model shows that the posterior shifts toward prior as BC decreases. **b** Preferred ITD of midbrain neurons doesn't shift toward zero as BC decreases. **c** Best ITD for perfectly correlation sounds (BC = 1) and decorrelated sounds (BC = 0.25 for intracellular, BC = 0.3 for extracellular). The *dashed line* is the regression line. **d** Predicted responses at the best ITD as BC varies needed to represent the posterior. **e** Measured responses at the best ITD as BC varies in the optic tectum (adapted from Saberi et al. 1998)

mechanisms. First, a shift of preferred directions toward zero degrees as binaural correlation decreases would be consistent with neural activity representing the posterior distribution. However, we found that the preferred ITD does not change as binaural correlation decreases (Fig. 4b and c). The best ITDs for the most peripheral neurons shifted slightly away from zero, rather than toward zero, when binaural correlation was reduced from 1 to 0.3 or 0.25 (Fig. 4c). The regression line fitting the relationship between the best ITD for perfectly correlated sounds and the best ITD at the low binaural correlation value had a slope of 0.97, indicating that best ITD did not change with binaural correlation (r^2=0.82). Supporting the claim that preferred ITD does not change with binaural correlation is the observation that neural responses to ITD in the owl's sound localization pathway are well described by a cross-correlation of the left and right input signals (Albeck and Konishi 1995; Saberi et al. 1998). As specified in the cross-correlation model, the preferred ITD does not change with binaural correlation, but the gain of the response decreases as binaural correlation decreases (Albeck and Konishi 1995; Saberi et al. 1998).

A second possible mechanism for representing the posterior in neural activity is for the gain of neural responses to change non-uniformly over the space map as binaural correlation decreases. Using the Bayesian model of Fischer and Peña (2011), we found that the pattern of gain change of individual neurons with binaural correlation at the preferred direction necessary to represent the posterior in neural activity is sigmoidal, with a greater diversity of responses across neurons at 0 binaural correlation than when binaural correlation is 1 (Fig. 4d). The greater diversity of responses across neurons at 0 binaural correlation occurs in the model because neurons with preferred directions in the periphery must all decrease their responses to zero while those neurons with preferred directions near zero must maintain a variety of elevated responses in order to encode a posterior that peaks near zero. This is in contrast to the experimentally observed ramp pattern where the gain converges to zero with little variability over neurons (Fig. 4e) (Albeck and Konishi 1995; Saberi et al. 1998). Therefore, the responses of midbrain auditory neurons in the owl are not consistent with the representation of the posterior probability directly in the population responses as binaural correlation changes.

3.3 Inference with a single spike

The owl's natural sound localization behavior requires inferences to be made rapidly about sound source direction. Here we test whether inference can be performed instantaneously in the non-uniform population code model. We consider the Bayesian model of ITD-based localization in the owl of Fischer and Peña (2011) in the case of single spikes per neuron. We use a small time interval so that the maximum

spike count per stimulus of the Poisson spiking neurons is approximately one (Fig. 5a). We compared the performance of the population vector decoder to the Bayesian estimator for different levels of binaural correlation. This allowed us to test whether Bayesian inference can be performed instantaneously in the model. This provides a stringent test of the model's applicability to the owl's natural behavior. It also allows us to test generally whether the non-uniform population code model applies to rapid inference problems. This is in contrast to the version of the sampling hypothesis that postulates that sampling over time is used for inference (Fiser et al. 2010).

We found that Bayesian inference can be implemented with a single spike per neuron in the non-uniform population code model. The population vector accurately approximates the performance of the Bayesian estimate of direction over a range of binaural correlation levels (Fig. 5b). Thus, the model is able to represent the stimulus uncertainty and produce the correct weighting of sensory and prior information when neurons fire at most one spike per neuron. The neural population encodes a decrease in binaural correlation by spreading activity over more of the population (Fig. 5a). This corresponds to a broadening of neural tuning as correlation decreases. Under the non-uniform population code model, the likelihood function is represented by the population activity. Thus, the population contains a representation of the stimulus uncertainty through this broadening of the activity pattern. While we ultimately measure the influence of the uncertainty on the direction estimate, the uncertainty is represented in the population and could be used in other computations.

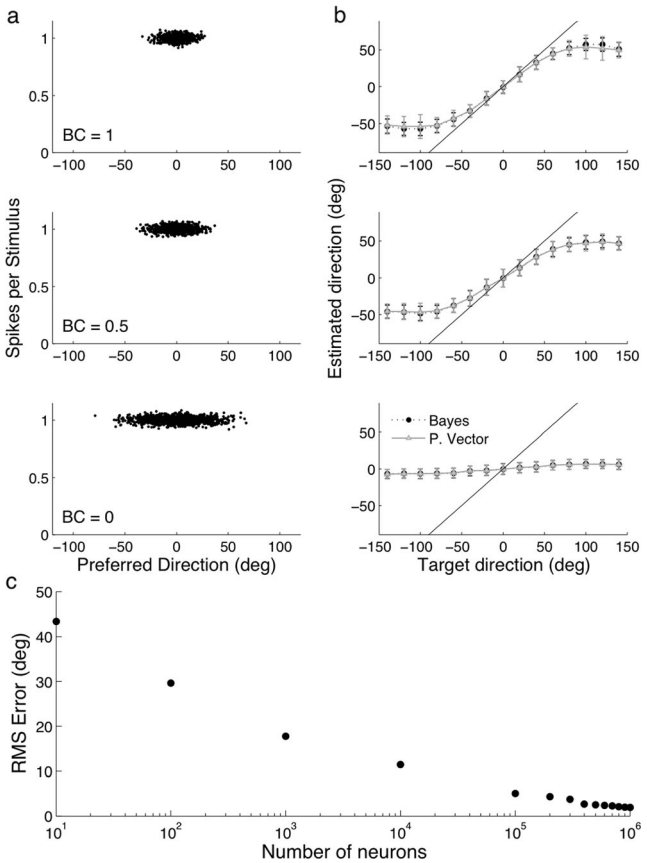

Fig. 5 Instantaneous inference **a** Responses of neurons in the model of Fischer and Peña (2011) for correlated sounds (*top*; BC = 1), decorrelated sounds (*middle*, BC = 0.5) and uncorrelated sounds (*bottom*; BC = 0). Jitter was added to the firing rates for display. **b** The population vector decoder matches the Bayesian estimate when single spikes per neuron are used. **c** Accurate neural inference with the population vector and single spikes per neuron requires a large population

While instantaneous inference is possible, the population vector only approximates the Bayesian estimate to within the owl's behavioral resolution of 3 deg (Bala et al. 2003) for a population size of more than 400,000 neurons (Fig. 5c). The midbrain auditory space map contains more than the required 400,000 neurons (Knudsen 1983). However, correlations in the responses of midbrain neurons due to shared inputs may limit the capacity for pooling to improve the accuracy of the population vector (Fischer and Peña 2011). Note that if multiple spikes are allowed, the number of required neurons decreases. The owl typically integrates information over 50–100 ms before generating an orienting movement (Knudsen et al. 1979), which would allow for several spikes per active neuron. This shows that, in principle, the non-uniform population model can implement Bayesian inference instantaneously if many independent neurons are used in the representation.

4 Discussion

Given the success of Bayesian models in describing perception and behavior, a central question is how a prior distribution can be represented and combined with sensory information in the brain. We tested several hypotheses for the neural representation of probability distributions in the owl's localization pathway, where the localization behavior for brief sounds is well described as Bayesian inference (Fischer and Peña 2011). We found that the owl's representation of auditory space is not consistent with the sampling hypothesis' prediction that stimulus-driven responses are samples from a posterior distribution and spontaneous responses are samples from the prior (Berkes et al. 2011; Fiser et al. 2010). We also found that neural responses in the owl's midbrain do not reflect a direct encoding of the posterior distribution in the population (Anderson and Van Essen 1994; Barber et al. 2003; Eliasmith and Anderson 2004; 1993; Gold and Shadlen 2000; Sahani and Dayan 2003; Sanger 1996; Simoncelli 2003, 2009). The responses are only consistent with the non-uniform population code model which proposes that the likelihood is represented in the shape of the tuning curves and the prior is represented in the distribution of preferred directions of midbrain neurons (Fischer and Peña 2011). In the non-

uniform population code model, variability in neural responses is treated as noise that is reduced by averaging the responses of many neurons (Fischer and Peña 2011). The increase in variability of spike count responses we observed as stimulus correlation increases is consistent with Poisson-like neural noise and is expected due to the increase in mean firing rate as correlation increases. In addition, our observation that best ITD does not change with binaural correlation matches the prediction of the non-uniform population code that the neural responses encode the likelihood function (Fischer and Peña 2011). Thus, the responses of midbrain auditory neurons are consistent with non-uniform population code model for the representation of probabilities.

There are additional models for the neural representation of probability distributions that we did not consider here. For example, the probabilistic population coding model proposes that neural activities are parameters in a high-dimensional probability distribution that is used for inference (Ma et al. 2006). This class of models has successfully described probabilistic computations in multiple settings (Beck et al. 2008; Fetsch et al. 2011; Ma et al. 2011). However, we have previously shown that this model does not apply to the owl's rapid sound localization behavior (Fischer and Peña 2011). We also did not test spiking models of inference that apply to binary variables (Deneve 2008) because sound source direction is a continuous variable.

What is the benefit of representing the prior in the distribution of preferred directions rather than directly representing the posterior in the neural activity? One explanation provided by Ganguli and Simoncelli (2014) is that such a representation produces an efficient code. This representation maximizes the information that the neural responses convey about the environmental variables. Additionally, if the posterior is directly encoded in the activities, then the prior needs to be supplied in terms of activity from either spontaneous responses or another population (Ma et al. 2006). This is inefficient in terms of the use of neural energy resources (Simoncelli 2009). A further benefit of the representation of the likelihood, rather than the posterior, in neural activity is that neurons can be pooled to reduce neural noise when an estimate is decoded from the population. The largest number of preferred directions, and therefore the largest reduction in neural noise, occurs at high-probability regions of the prior. Therefore, representing the prior in the distribution of preferred directions can have the benefit of providing a more accurate neural approximation to the Bayesian estimate.

The prediction of the sampling hypothesis that the distribution of spontaneous activity of a neuron is the average of the distribution of the stimulus driven activity is independent of the particular features of the Bayesian model. Therefore, our finding that the prediction does not hold cannot be explained by a failure of our particular Bayesian model of localization in the owl. Berkes et al. (2011) found that the prediction of the sampling hypothesis holds in the ferret primary visual cortex for natural stimuli, but not for artificial stimuli. The natural stimuli for the owl are rustling sounds or vocalizations of prey or other owls (Konishi 1973). The noise bursts used in this study are approximations to the short, broadband rustling sounds of prey (Konishi 1973). In addition, noise bursts readily elicit orienting behavior in owls (Knudsen et al. 1979). Even though the stimuli are approximations to natural stimuli, the lack of spontaneous activity in the auditory midbrain neurons indicates that the average of stimulus-driven activity from any stimulus distribution will not match the spontaneous activity. The stimulus level used to measure driven responses will influence the match between the average of stimulus-driven activity and the spontaneous activity. As the stimulus level decreases, the match will improve (Arthur 2004). However, for any supra-threshold stimulus level, the average of stimulus-driven activity and the spontaneous activity will never be equal, as required by the sampling hypothesis. High spontaneous activity is observed in the superficial layers of the optic tectum (Knudsen 1982), which may provide a means to represent prior information. However, these neurons do not have auditory evoked responses and thus their responses would not correspond to sampling from the posterior, as required in the sampling hypothesis. Finally, our rejection of the sampling hypothesis in the owl is supported by the observation that response variability increases when sensory noise decreases. We found that this was true for preferred stimulus parameters that generated the highest firing rate and for non-preferred stimulus parameters where the neuron did not spike (Fig. 3). This shows that even for stimuli that produce average responses that match the spontaneous activity, the variability is not consistent with the prediction of the sampling hypothesis. We also note that an alternative version of the sampling hypothesis could be implemented in a population of neurons where sampling occurs over neurons, as well as over time. Sampling over neurons would have the benefit of allowing for rapid inference if many neurons represented each feature. However, increasing the number of neurons per feature-dimension of the distribution would limit the possibility of representing high-dimensional distributions with small numbers of neurons. While sampling over neurons and time may be possible in general, this model would lead to the same predictions for neural responses that we considered here and would therefore be inconsistent with ICx responses.

There is a trade-off between the time and number of neurons required to do inference in the sampling hypothesis and non-uniform population code models. The sampling hypothesis, as presented in Fiser et al. (2010) and Berkes et al. (2011), requires only a single neuron per dimension of the distribution, but it takes time to generate the samples necessary to represent the distribution. In a sampling scheme, a sample from the distribution is generated at each time, but many such samples are required to perform inferences such as computing

the mean, variance, or higher order statistics of the distribution accurately. This is advantageous for representing the high-dimensional distributions that are likely involved in complex cognitive tasks. By contrast, the non-uniform population code can perform inference with a single spike per neuron, but requires many neurons to represent the distribution. This is advantageous for tasks, such as orientation that must be performed rapidly and are relatively stereotyped. We propose that the method used to represent Bayesian models in the brain may vary among regions depending on the task demands.

Acknowledgments We thank W. Cox for comments on the manuscript. This work was funded by U.S. National Institutes of Health grant DC012949.

Conflict of interest The authors declare that they have no conflict of interest.

References

Albeck, Y., & Konishi, M. (1995). Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals. *Journal of Neurophysiology, 74*, 1689–1700.

Anderson, C.H., and Van Essen, D.C. (1994). Neurobiological computational systems. *Computational Intelligence Imitating Life*, 213–222.

Arthur, B. J. (2004). Sensitivity to spectral interaural intensity difference cues in space-specific neurons of the barn owl. *Journal of Comparative Physiology A, 190*, 91–104.

Bala, A. D. S., Spitzer, M. W., & Takahashi, T. T. (2003). Prediction of auditory spatial acuity from neural images on the owl's auditory space map. *Nature, 424*, 771–774.

Barber, M. J., Clark, J. W., & Anderson, C. H. (2003). Neural representation of probabilistic information. *Neural Computation, 15*, 1843–1864.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E., & Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron, 60*, 1142–1152.

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science, 331*, 83–87.

Deneve, S. (2008). Bayesian spiking neurons I: inference. *Neural Computation, 20*, 91–117.

Edut, S., & Eilam, D. (2004). Protean behavior under barn-owl attack: voles alternate between freezing and fleeing and spiny mice flee in alternating patterns. *Behavioural Brain Research, 155*, 207–216.

Eliasmith, C., & Anderson, C. C. H. (2004). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. Cambridge: MIT Press.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience, 15*, 146–154.

Fischer, B. J., & Peña, J. L. (2011). Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience, 14*, 1061–1066.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Science, 14*, 119–130.

Foldiak, P. (1993). The "Ideal Homunculus": statistical inference from neural population responses. In Eeckman, F. & Bower, J. (eds), *Computation and Neural Systems* (pp. 55–60). Kluwer Academic Publishers.

Ganguli, D., & Simoncelli, E. P. (2014). Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Comput, 26*, 2103–2134.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*, 926–932.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404*, 390–394.

Hausmann, L., von Campenhausen, M., Endler, F., Singheiser, M., & Wagner, H. (2009). Improvements of sound localization abilities by the facial ruff of the barn owl (Tyto alba) as demonstrated by virtual ruff removal. *PLoS One, 4*, e7721.

Kita, H., & Armstrong, W. (1991). A biotin-containing compound N-(2-aminoethyl)biotinamide for intracellular labeling and neuronal tracing studies: comparison with biocytin. *Journal of Neuroscience Methods, 37*, 141–150.

Knudsen, E. I. (1982). Auditory and visual maps of space in the optic tectum of the owl. *Journal of Neuroscience, 2*, 1177–1194.

Knudsen, E. I. (1983). Subdivisions of the inferior colliculus in the barn owl (Tyto alba). *Journal of Comparative Neurology, 218*, 174–186.

Knudsen, E. I., Blasdel, G. G., & Konishi, M. (1979). Sound localization by the barn owl (Tyto alba) measured with the search coil technique. *Journal of Comparative Physiology, 133*, 1–11.

Konishi, M. (1973). How the owl tracks its prey: experiments with trained barn owls reveal how their acute sense of hearing enables them to catch prey in the dark. *American Scientist, 61*, 414–424.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*, 1432–1438.

Ma, W. J., Navalpakkam, V., Beck, J. M., van den Berg, R., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience, 14*, 783–790.

Peña, J. L., & Konishi, M. (2002). From postsynaptic potentials to spikes in the genesis of auditory spatial receptive fields. *Journal of Neuroscience, 22*, 5652–5658.

Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B. J., & Farahbod, H. (1998). Effects of interaural decorrelation on neural and behavioral detection of spatial cues. *Neuron, 21*, 789–798.

Sahani, M., & Dayan, P. (2003). Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Computation, 15*, 2255–2279.

Salinas, E. (2011). Prior and prejudice. *Nature Neuroscience, 14*, 943–945.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology, 76*, 2790–2793.

Shi, L., and Griffiths, T.L. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. In *Advances in neural information processing systems*, pp. 1669–1677.

Simoncelli, E.P. (2003). Local analysis of visual motion. *Visual Neuroscience*, 1616–1623.

Simoncelli, E.P. (2009). Optimal estimation in sensory systems. *Cognitive Neuroscience*, IV.

Wang, Y., Shanbhag, S. J., Fischer, B. J., & Peña, J. L. (2012). Population-wide bias of surround suppression in auditory spatial receptive fields of the owl's midbrain. *Journal of Neuroscience, 32*, 10470–10478.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*, 598–604.