

# Bayesian Inference with Efficient Neural Population Codes

Xue-Xin Wei and Alan A. Stocker\*

Department of Psychology, University of Pennsylvania, U.S.A.  
weixxpku@gmail.com, astocker@sas.upenn.edu

**Abstract.** The accuracy with which the brain can infer the value of a stimulus variable depends on both the amount of stimulus information that is represented in sensory neurons (encoding) and the mechanism by which this information is subsequently retrieved from the responses of these neurons (decoding). Previous studies have mainly focused on either the encoding or the decoding aspect. Here, we present a new framework that functionally links the two. More specifically, we demonstrate that optimal (efficient) population codes which guarantee uniform firing rate distributions allow the accurate emulation of optimal (Bayesian) inference using a biophysically plausible neural mechanism. The framework provides predictions for estimation bias and variability as a function of stimulus prior, strength and integration time, as well as physiological parameters such as tuning curves and spontaneous firing rates. Our framework represents an example of the duality between representation and computation in neural information processing.

**Keywords:** Neural encoding, Optimal decoding, Population vector.

## 1 Introduction

Our world is uncertain - the sensory information we have access to is typically limited, ambiguous, and noisy. The goal of perception is to infer the state of the world from this unreliable information. In that sense, our percept of the world remains an *educated guess* at best. An increasing number of psychophysical and modeling studies suggest, however, that the brain's "guessing" is often statistically optimal and well described by Bayesian inference. This has been shown for perceptual tasks such as stimulus detection (*e.g.* [1]), discrimination (*e.g.* speed [2], orientation [3], contrast [4]), and integration (*e.g.* [5]). The Bayesian framework not only provides an account for how the brain might compute with noisy sensory information, but also how it should combine this information with accumulated prior knowledge in an optimal, probabilistic way.

Some efforts have recently been allocated to the understanding of the neural mechanisms underlying probabilistic inference. A major difficulty in achieving

---

\* We thank Josh Gold, David Brainard, and Alex Tank for helpful comments on the manuscript. Supported by grant ONR-N000141110744.

such understanding is the fact that neural computation does not allow a clear distinction between representation and computation. In order to emulate Bayesian inference the brain has to adopt a way of representing prior beliefs and likelihood functions, and performing the inference process using the *same* neural substrate. Some proposals suggest how the brain can represent and combine likelihood functions in light of the intrinsic variability in neural responses [6,7], but do not specify how to incorporate prior beliefs or how to perform inference. Other proposals suggest that prior beliefs are embedded in the distribution of tuning curves (with regard to the stimulus space) within a neural population [8]. Furthermore, it has been shown that efficient neural codes could predict both, neural tuning characteristics in sensory areas as well as psychophysically measured discrimination thresholds [10]. Finally, a recent study demonstrates that Bayesian inference can be approximated under the assumption that the shape of the tuning curves of the neurons in a population directly reflect the likelihood function [11].

We propose a model framework on how the brain emulates Bayesian inference, which combines and completes some of the concepts listed above. Specifically, we demonstrate that efficient coding leads to a neural representation for which a population vector with a point nonlinearity provides a good approximation of a Bayes' least squared-error estimate. We validate our framework with simulations showing how well it can approximate the explicit Bayesian decoder, and how its decoding performance compares to other popular decoding methods.

## 2 Model Framework

Let us consider the following perceptual inference problem: Suppose a population of  $N$  sensory neurons is tuned to a stimulus variable  $\theta$  such that a specific value of  $\theta$  elicits a specific population response characterized as the vector  $R = [r_1, r_2, \dots, r_N]$ , where  $r_i$  is the spike-count over time  $\tau$  of the  $i_{th}$  neuron. Inherent stochastic variability in the spike-counts causes  $R$  to be a noisy representation of the stimulus variable. The task is then to derive an estimate  $\hat{\theta}$  of the true value of  $\theta$  from  $R$ .

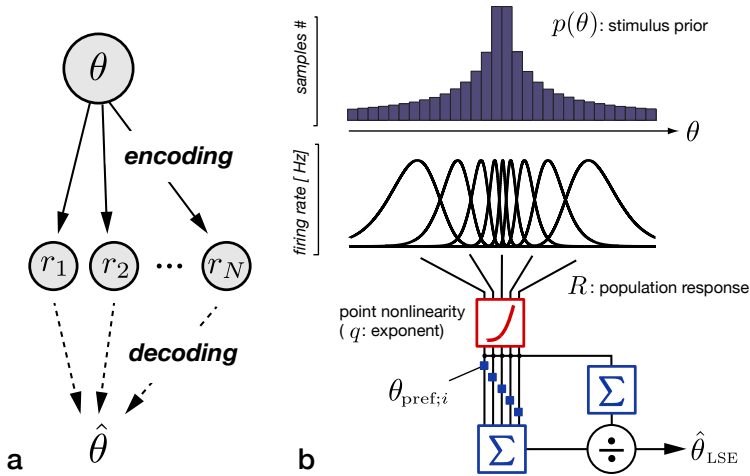
We can consider the estimation problem as probabilistic inference over the simple directed graph shown in Fig. 1a: Encoding is the generative process that characterizes the dependency of the population response  $R$  on the stimulus variable  $\theta$ , while decoding is the inverse process of defining an estimate  $\hat{\theta}$  as a function of the population response  $R$ <sup>1</sup>.

### 2.1 Optimal (Bayesian) Decoder

We first formulate the explicit Bayesian decoder, which will serve us as a benchmark for assessing decoding performance. With the assumption that the variability in individual neurons' spike-counts is governed by independent Poisson

---

<sup>1</sup> In the context of this paper, we consider *inferring*, *decoding*, and *estimating* as synonymous.



**Fig. 1.** *Neural emulation of Bayesian inference.* a) Generative model describing the encoding of a stimulus value  $\theta$  in the response  $R = [r_1, r_2, \dots, r_N]$  of a population of neurons. Decoding is the process of inferring an estimate  $\hat{\theta}$  of the stimulus value from  $R$ . b) The proposed encoding-decoding framework for computing the least squared-error estimate  $\hat{\theta}_{\text{LSE}}$ : a neural population that efficiently encodes the stimulus according to the stimulus prior  $p(\theta)$ , followed by a *generalized population vector decoder* with a power-law nonlinearity with exponent  $q$  (see Eq. (6)).

processes, we write the likelihood function over  $\theta$  (for a given integration time  $\tau$ ) as

$$p(R|\theta) = \prod_{i=1}^N \frac{(\tau f_i(\theta))^{r_i}}{r_i!} e^{-\tau f_i(\theta)}, \quad (1)$$

with  $f_i(\theta)$  describing the expected firing rate of neuron  $i$  as a function of  $\theta$  (tuning curve). We define the optimal estimator  $\hat{\theta}_{\text{LSE}}$  as the estimator that minimizes the expected squared-error between the estimate and the true stimulus value, thus

$$\hat{\theta}_{\text{LSE}}(R) = \frac{\int \theta p(R|\theta) p(\theta) d\theta}{\int p(R|\theta) p(\theta) d\theta}, \quad (2)$$

using Bayes' rule to appropriately combine the sensory evidence with the stimulus prior distribution  $p(\theta)$ .

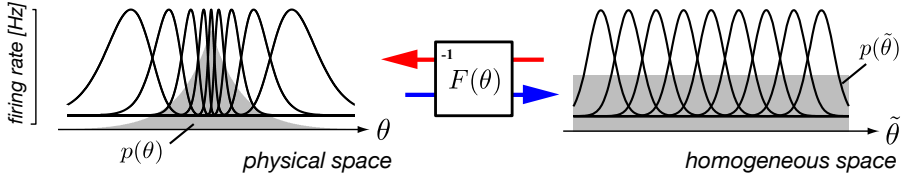
Note, that given our choice of loss function, Eq. (2) uniquely defines the optimal decoder for arbitrary stimulus priors and tuning curve distributions and shapes. In the following we will demonstrate how efficiently encoded stimulus information allows us to approximate the optimal Bayesian decoder with the generalized population vector read-out mechanism shown in Fig. 1b.

### 2.2 Optimal (Efficient) Encoding

Efficient encoding proposes that the tuning characteristics of a neural population are adapted to the stimulus prior distribution  $p(\theta)$  such that the population optimally represents the stimulus [9]. Different definitions of “efficient” are possible. In the following, we adopt the definition of an efficient neural code for which the *total expected spike count* of each neuron in the population is the same, thus a homeostatic code that equally employs all the neurons in the population.

In order to create such code, we first map the physical space with units  $\theta$  to an equivalent *homogeneous space* with units  $\tilde{\theta} = F(\theta)$  in which the stimulus prior distribution is flat. This defines the mapping as  $F(\theta) = \int^\theta p(\chi)d\chi$ , thus as the cumulative of the prior distribution  $p(\theta)$  (*i.e.* histogram equalization). We then assume a population of neurons with identical tuning curves that uniformly tiles this homogeneous space. Such population provides an efficient encoding of the stimulus according to above definition. Interestingly, it also maximizes mutual information between stimulus and the population’s response for given homogeneous tuning width and noise characteristics [10].

The tuning curves in the physical space are simply obtained by applying the inverse transformation  $\theta = F^{-1}(\tilde{\theta})$  to the homogeneous population. Figure 2 illustrates the tuning curves of an efficient neural population and the mapping to/from the homogeneous space for an exponential prior distribution  $p(\theta)$  (shaded area).



**Fig. 2.** Efficient encoding and the definition of the homogeneous space

The definition of the homogeneous space allows us to formulate two crucial approximations with regard to the optimal decoding of the population: First, it guarantees that the distribution of neurons’ preferred tuning values  $\theta_{\text{pref}}$  represents an accurate sample of the prior density  $p(\theta)$ . Consequently, we can approximate the expected value of the prior density by the average preferred tuning of the population, thus

$$\frac{\int \theta p(\theta)d\theta}{\int p(\theta)d\theta} \approx \frac{\sum_i \theta_{\text{pref},i}}{\sum_i 1} . \tag{3}$$

Second, because the likelihood is a function over  $\theta$  and not a probability density, we can compute its values directly in the homogeneous space. With the assumption that the tuning curves  $\tilde{f}_i(\tilde{\theta})$  in the homogeneous space are Gaussians of

identical width and amplitude<sup>2</sup> we can approximate the likelihood function Eq. (1) as

$$\begin{aligned}
 p(R|\tilde{\theta}) &= \prod_i \frac{(\tau \tilde{f}_i(\tilde{\theta}))^{r_i}}{r_i!} e^{-\tau \tilde{f}_i(\tilde{\theta})} \approx c_1 \prod_i \tilde{f}_i(\tilde{\theta})^{r_i} \\
 &\approx c_1 e^{(r_1+r_2+\dots+r_N)(\tilde{\theta}-\mu)^2/2\sigma^2}, \quad (4)
 \end{aligned}$$

where  $c_1$  is a constant,  $\sigma$  is the width of the homogeneous Gaussian tuning curve, and  $\mu = \sum_i \tilde{\theta}_{\text{pref};i} r_i / \sum_i r_i$ . The approximate likelihood has the same shape as the homogeneous tuning curve (Gaussian) *scaled* by the exponent  $q = \sum_i r_i$ , which is the total spike-count. In addition, the population response  $R$  itself is a noisy sample of the homogeneous tuning curve, approximately centered around its empirical mean  $\mu$ . Taken together, we can treat the population response as a scaled and noisy representation of the likelihood function, sampled at the preferred tuning  $\theta_{\text{pref};i}$  of each neuron. This leads to the approximation

$$r_i^q \approx c_2 p(R|\theta_{\text{pref};i}), \quad (5)$$

where  $c_2$  is a global constant.

### 2.3 Generalized Population Vector Decoding

Now, let us revisit the optimal Bayes least squared-error decoder Eq. (2) in the context of above approximations. We can now replace the expressions for the likelihood and the integral over the prior distribution with Eqs. (3) and (5), respectively, and find

$$\hat{\theta}_{\text{LSE}} = \frac{\int \theta p(\theta) p(R|\theta) d\theta}{\int p(\theta) p(R|\theta) d\theta} \approx \frac{\sum_i \theta_{\text{pref};i} r_i^q}{\sum_i r_i^q} \quad (6)$$

as a valid approximation. We notice that the right-hand side resembles the familiar population vector decoder [12], modified by a power-law nonlinearity on the spike-counts  $r_i$ . Thus, Bayesian inference can be approximated by a generalized population vector decoder!

Fig. 1b illustrates a functional diagram of the decoder. It shares the same biophysically plausible basic operations of the population vector decoder namely taking a weighted sum over the population spike-counts followed by divisive normalization. The proposed modification only involves a point nonlinearity that can be thought of taking place *e.g.* in dendritic processes of the decoder stage.

<sup>2</sup> Given our constraints, a truly homogeneous space would need to be compact (circular). In this space, a similar analysis using von Mises tuning curves results in an exponent

$$q = \sqrt{\sum_i \sum_j r_i r_j \cos(\theta_i - \theta_j)}$$

where  $\theta_i$ 's are the modes of the von Mises tuning curves.

**The Magnitude of Exponent  $q$ .** The above derivation of the optimal exponent  $q$  was based on the assumption that neurons in the population do not show any spontaneous activity. Assuming a more realistic neuron model that includes spontaneous activity  $b$ , the tuning curve (in the homogeneous space) can be written as  $\tilde{f}_i(\tilde{\theta}) = \tilde{f}_{\text{evoked}}(\tilde{\theta}) + b$ , which leads to a likelihood function that is no longer strictly Gaussian (*i.e.* von Mises, respectively). In this case, a second-order Taylor expansion allows us to derive an approximate description for  $q$  that we do not show here in detail. Generally, we find that spontaneous activity as well as noise correlations significantly reduces the magnitude of  $q$  while the specific values depend on the ratio between the width (spread) of the population response and the amount of stimulus information contained. Spontaneous activity and noise correlations decreases the amount of information while the spread of the population response is given by the tuning curves.

### 3 Results

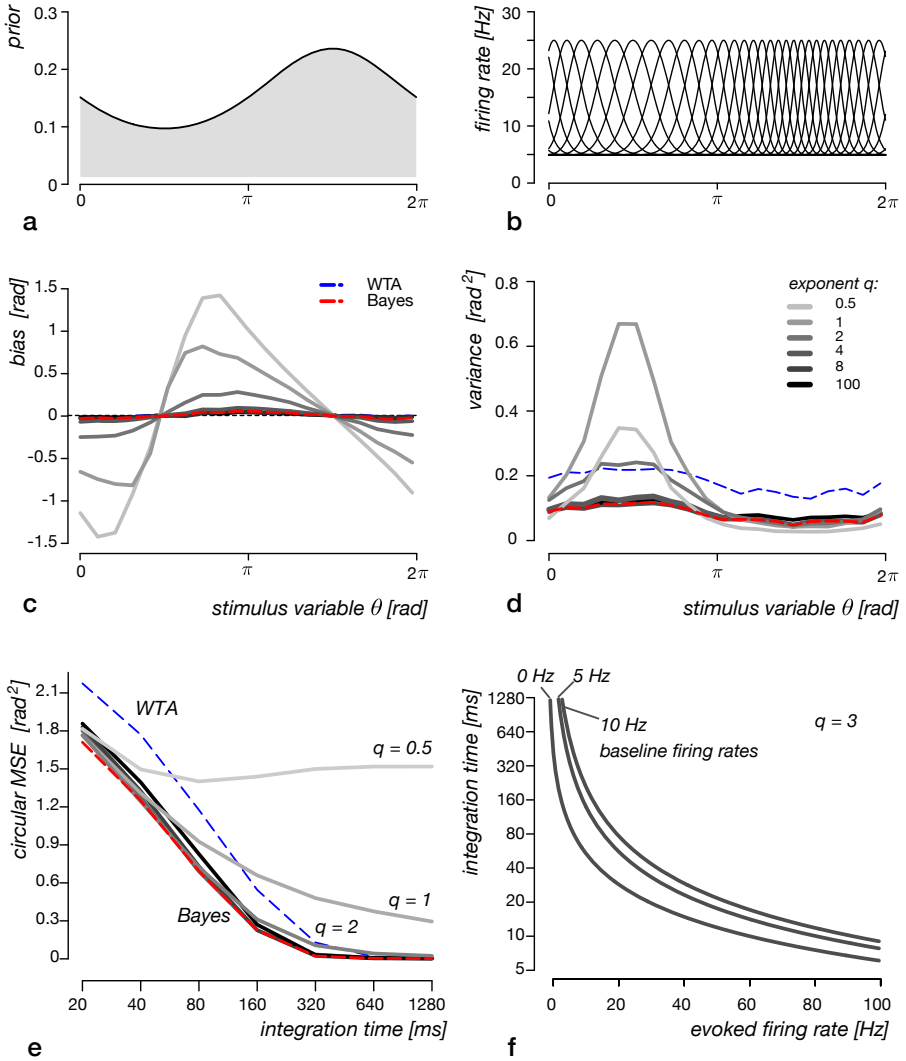
We performed numerical simulations to validate our new encoding-decoding framework (see Fig. 3). Choosing an arbitrary stimulus prior  $p(\theta)$ , we first created an efficient neural representation with a population of  $N = 30$  neurons with von Mises tuning curves (in the homogeneous space  $\tilde{\theta}$ ), peak-evoked firing rates of 20 Hz, and spontaneous activity of 5 Hz. We then computed estimation bias and variance for each stimulus value  $\theta$  over 2000 samples of the population response  $R(\theta)$  (*i.e.* spike-counts in a  $\tau = 160$  ms time interval) for i) the explicit Bayesian decoder, ii) the WTA decoder, iii) and the generalize population vector decoder with a range of different exponents  $q \in [0.5, 100]$ .

As expected, the explicit Bayesian estimator shows a bias toward the peak of the prior as well as minimal/maximal variance at the peak/trough of the prior. While the generalized population vector qualitatively exhibits the same overall behavior, only exponents  $q \geq 3$  lead to a good quantitative match<sup>3</sup>.

Figure 3e illustrates the decoding performance measured as the mean squared-error (MSE) over the stimulus prior. Performance is plotted as a function of integration time  $\tau$ . On one hand, the traditional population vector ( $q = 1$ ) performs well for short  $\tau$  while larger exponents are necessary for a good approximation at longer  $\tau$ . A *fixed* intermediate exponent  $q \in [3, 4]$ , however, leads to a good approximation of the Bayesian decoder over the entire range of integration times. Interestingly, such values in particular, and the power-law nonlinearity in general seem to be supported by physiologically measured nonlinearities in neural signal transmission [13].

Finally, a fixed  $q$  can provide optimal performance if the integration time is adjusted accordingly, depending on the strength of the evoked response (see Fig. 3f). Thus, while the known change in integration time of biological sensory systems is considered to reflect a speed-accuracy trade-off, it might also serve to keep decoding at its optimal operating point.

<sup>3</sup> Smoothing the population response with a small convolutional kernel  $K$  leads to a slightly better performance, without, however, significantly changing the results.



**Fig. 3.** Estimation performance. a) Stimulus prior. b) Efficient population encoding (von Mises tuning curves). c,d) Estimation bias and variance for the explicit Bayesian solution, the WTA read-out, and our generalized population vector decoder for different exponents  $q$ . e) Circular MSE for different values of the exponent  $q$ . f) A fixed exponent  $q$  can approximate the full Bayesian estimator if the integration time is adjusted according to stimulus strength (evoked activity), with weaker stimuli requiring longer integration times.

## 4 Conclusions

Understanding the neural mechanisms by which the brain actually solves probabilistic inference tasks remains a challenging question. We present a solution that addresses both the encoding and decoding of sensory information. We show that if encoding is efficient, then inference can be emulated with a biophysically plausible read-out mechanism that is similar to the well-known population vector decoder<sup>4</sup>. In fact, the framework defines an entire class of decoding mechanisms that includes the standard population vector read-out and the WTA decoder.

Efficient encoding of sensory information in the specific form proposed here is at the heart of the framework. It guarantees both that the likelihood is well approximated by the scaled population response and that the prior is embedded in the distribution of neural tuning across the stimulus space.

## References

1. Burgess, A.: Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *Journal of Optical Society of America A* 2(9), 1498 (1985)
2. Stocker, A.A., Simoncelli, E.P.: Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience* 9(4), 578–585 (2006)
3. Girshick, A.R., Landy, M.S., Simoncelli, E.P.: Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience* 14(7), 926–932 (2011)
4. Sanborn, A.N., Dayan, P.: Optimal decisions for contrast discrimination. *Journal of Vision* 11(14) (2011)
5. Knill, D.: Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision* 7(7), 1–24 (2007)
6. Jazayeri, M., Movshon, J.A.: Optimal representation of sensory information by neural populations. *Nature Neuroscience* 9, 690 (2006)
7. Ma, W.J., Beck, J.M., Latham, P.E., Pouget, A.: Bayesian inference with probabilistic population codes. *Nature Neuroscience* 9, 1432 (2006)
8. Simoncelli, E.P.: Optimal Estimation in Sensory Systems. In: *The New Cognitive Neuroscience*, 4th edn. MIT Press (2009)
9. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. In: Rosenblith, W.A. (ed.) *Sensory Communication*, pp. 217–234. MIT Press, Cambridge (1961)
10. Ganguli, D., Simoncelli, E.P.: Implicit encoding of prior probabilities in optimal neural populations. In: *Adv. Neural Information Processing Systems NIPS*, vol. 23, pp. 658–666 (2011)
11. Fischer, B.J.: Bayesian estimates from heterogeneous population codes. In: *Proc. IEEE Int’l. Joint Conf. on Neural Networks. IEEE* (2010)
12. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: Neuronal population coding of movement direction. *Science* 233(4771), 1416 (1986)
13. Miller, K.D., Troyer, T.W.: Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.* 7(2), 653–659 (2002)

---

<sup>4</sup> It just came to our attention that a related albeit different framework has been recently suggested (D.Ganguli and E.Simoncelli, Abstract presented at CoSyNe 2012).