A neural network with a biologically possible architecture can implement Bayesian estimation and reproduce Piéron’s law

Daiki Futagi\(^{1a)}\) and Katsunori Kitano\(^{2}\)

\(^{1}\) Graduate School of Information Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 5258577, Japan

\(^{2}\) Department of Human and Computer Intelligence, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 5258577, Japan

\(^{a)}\) futagi@cns.ci.ritsumei.ac.jp

Received August 15, 2015; Revised November 10, 2015; Published April 1, 2016

Abstract: Bayesian estimation theory has been expected to explain how the brain deals with uncertainty. Several previous studies have implied that cortical network models could implement Bayesian computation. However, the feasibility of the required computational procedures is still unclear under the physiological and anatomical constraints of neural systems. Here, we propose a neural network model that implements the algorithm in a biologically realizable manner, incorporating discrete choice theory. Our model successfully demonstrates an orientation discrimination task with significantly noisy visual images and the relation between the stimulus intensity and the reaction time known as Piéron’s law.

Key Words: Bayesian estimation theory, neural network, discrete choice method, normalization, Piéron’s law

1. Introduction

In order to adapt to the uncertain external world, the brain utilizes prior knowledge in conjunction with evidence obtained through repeated observation [1, 2]. Bayesian theory inspired by such a brain function provides a mathematical framework for the cognitive process. In recent years, an analysis based on Bayesian theory provided a novel interpretation of how prior knowledge and sensory information are connected in perception and what kind of information neural activity encodes [3]. Furthermore, it is shown that a commonly used model of the cerebral cortex can implement Bayesian estimation for an arbitrary hidden Markov model [4]. In the previously proposed model, the firing rates of a neural population encode the probability distribution of the hidden states of a Markov model. After repeated observations, the neural population can represent the posterior distribution over the hidden states and correctly estimate the probability of the state, even in the presence of...
significant noise. However, it is still unclear whether the computational procedures in the model can be implemented under the biological constraints of the neural system. In particular, it is unclear how a neural system solves such a mathematical equation, even though the model suggests that the synaptic weights should satisfy a matrix equation of the relation between the Bayesian estimation algorithm and the model architecture. In this study, we propose a neural circuit model that performs such computational procedures of the Bayesian estimation algorithm in a biologically plausible manner. Our model is based on the previously proposed model but is inspired by the discrete choice (DC) method [5]. The DC method with synaptic noise obeying a long-tailed distribution (Gumbel distribution) plays an important role in the calculation of the normalization factor through the firing-rate dynamics of the neural networks. We investigate the performance of our model from a cognitive viewpoint such as the relation between the required estimation time and the stimulus intensity.

2. Method

2.1 Correspondence of the neural network dynamics with Bayesian computation

In the previous study [4], Bayesian estimation was implemented by a recurrent network that consists of neurons with firing-rate dynamics. The following equation describes the dynamics of the network:

$$\tau \frac{dV}{dt} = -V + WI + MV,$$

where \(\tau\) is a time constant, and \(I\) and \(V\) denote the input firing rates to the network and output firing rates of the neurons in the network, respectively. In addition, \(W\) and \(M\) represent the feedforward and recurrent synaptic weight matrices, respectively. This equation can be written in terms of the discrete-time dynamics as follows:

$$V_t (t + 1) = V_t (t) + \alpha \left( -V_t (t) + w_i I_t (t) + \sum_j M_{ij} V_j (t) \right)$$

$$= \alpha w_i I_t (t) + \sum_j m_{ij} V_j (t),$$

where \(\alpha\) is an integration rate, \(V_t\) is the \(i\)-th component of the vector \(V\), \(w_i\) is the \(i\)-th row of the matrix \(W\), and \(m_{ij} = \alpha M_{ij}\) for \(i \neq j\) and \(m_{ii} = 1 + \alpha (M_{ii} - 1)\), where \(M_{ij}\) is the element of \(M\) in the \(i\)-th row and the \(j\)-th column. In the model, a neuron encodes the probability of a state of the external world. Bayesian estimation in a simple hidden Markov model is described as

$$\ln \mathbb{P}(\theta_t | I(t),...,I(1)) = -\ln z(t) + \ln \mathbb{P}(I(t) | \theta_t^i)$$

$$+ \ln \left[ \sum_j \mathbb{P}(\theta_t^i | \theta_{t-1}^j) \mathbb{P}(\theta_{t-1}^j | I(t-1),...,I(1)) \right],$$

where \(\theta_t\) is one of the hidden states of the hidden Markov model, and \(\theta^t\) and \(I(t)\) represent the hidden state and observable input at time \(t\) respectively. In addition, \(z(t)\) is the normalization constant that is needed to maintain the equality in \(\sum_i \mathbb{P}(\theta_t^i | I(t),...,I(1))\). A recurrent network governed by Eq. (2) can implement Eq. (3) if

$$V_t (t + 1) = \ln \mathbb{P}(\theta_t^i | I(t),...,I(1)),$$

$$\alpha w_i I_t (t) = \ln \mathbb{P}(I(t) | \theta_t^i),$$

$$\sum_j m_{ij} V_j (t) = \ln \left[ \sum_j \mathbb{P}(\theta_t^i | \theta_{t-1}^j) \mathbb{P}(\theta_{t-1}^j | I(t-1),...,I(1)) \right],$$

with the logarithm normalization factor \(\ln \sum_i \mathbb{P}(\theta_t^i, I(t) | I(t-1),...,I(1))\). The normalization is added after updating the network state with Eq. (2). Therefore, Eq. (2) is rewritten as follows:
\[ V_i(t+1) = \alpha w_i I(t) + \sum_j m_{ij} V_j(t) - \ln \sum_j \exp (v_j(t)), \] (7)

\[ v_i(t) = \alpha w_i I(t) + \sum_j m_{ij} V_j(t). \]

In order to maintain the correspondence relation between the neural network dynamics and the Bayesian computation, \( m_{ij} \) is determined so that it should satisfy Eq. (6), and the normalization factor \( \ln z(t) \), which stabilizes the performance, takes the complicated form of the logarithm of a sum of a probability. However, it seems nontrivial whether these terms can be computed in a biologically plausible manner. Here, we focused on how the normalization factor can be computed on the basis of neurophysiological knowledge. For simplicity, we omitted the problem related to the estimation of the recurrent weight matrix \( M \); the recurrent weight \( m_{ij} = 0 \) for \( i \neq j \) and \( m_{ii} = 1 \). This means that the hidden external state is assumed to not be variable. Therefore, Eq. (7) is rewritten as follows:

\[ V_i(t+1) = \alpha w_i I(t) + V_i(t) - \ln \sum_j \exp (v_j(t)), \] (8)

\[ v_i(t) = \alpha w_i I(t) + V_i(t). \]

2.2 Estimation algorithm for the logarithm normalization factor

We utilize part of the DC method to estimate the logarithm of the normalization factor \( \ln \sum_j \exp(v_j(t)) \) in Eq. (8) simply [5]. The DC method is often used in marketing science to analyze one’s decision-making. In the model, it is assumed that a decision maker chooses one of several discrete choices depending on the utility of these choices. The utility \( U \) of the \( i \)-th choice is defined as

\[ U_i = u_i + \epsilon_i, \quad i = 1, ..., N, \] (9)

where \( u \) denotes the observable and nonrandom utility, and \( \epsilon \) denotes the unobservable and random utility. If \( \epsilon \) obeys the Gumbel distribution, i.e.,

\[ \epsilon \sim f(x; \mu, \eta) = \eta \exp (-\eta (x - \mu)) \exp (-\exp (-\eta (x - \mu))), \] (10)

then the expectation of the utility \( U \) is described as follows:

\[ E\left[ \max_i (U_i) \right] = \int_{-\infty}^{\infty} xf(x; S, \eta) dx \]
\[ = S + \frac{\gamma}{\eta} \]
\[ = \frac{1}{\eta} \ln \sum_{i=1}^{N} \exp \left( \eta \left( u_i + \mu + \frac{\gamma}{\eta} \right) \right), \] (11)

where \( S = \frac{1}{\eta} \ln \sum_{i=1}^{N} \exp (\eta (u_i + \mu)), \) and \( \gamma \) is Euler's constant. If the random variable \( \epsilon \) follows a particular Gumbel distribution \( f(x; -\gamma, 1) \) and the nonrandom variable \( u \) is regarded as \( v \) in Eq. (8), then the logarithm of the normalization factor in Eq. (8) is obtained as the expectation of the maximum composite value:

\[ E\left[ \max_i (v_i + \epsilon_i) \right] = \ln \sum_{i=1}^{N} \exp (v_i). \] (12)

2.3 Neuronal architecture of the proposed model

As shown in Eq. (12), the logarithm of the normalization factor in Eq. (8) could be obtained as the expectation of the maximum composite value by utilizing the properties of the Gumbel distribution. Thus, we extend the previous model in Eq. (8) with the supportive network model that approximately calculates the expectation value by repeated random sampling, i.e., a Monte Carlo method. The
proposed model consists of two layers, as shown in Fig. 1. Through the interaction between the two layers, the model computes the posterior probabilities including the normalization of the probability distribution in a way inspired by the DC method. In the following, we describe the mathematical model and its role in each layer in more detail.

![Network architecture of the proposed model](image)

**Fig. 1.** Network architecture of the proposed model. The circles labeled with “inh” and “exc” represent inhibitory and excitatory neurons, respectively, and the arrow represents the output destination of the neuron. The network has a columnar structure (shaded gray), and it is roughly divided into two layers: Layer A and Layer B (shaded red and green, respectively). Layer A is based on the previous model [4] that implements Bayesian computation. In order to maintain the correspondence of the network dynamics in Layer A and the Bayesian computation, Layer B implements the computation for the normalization. The neural groups in Layer B carry out the computation in parallel, each of which performs as a winner-take-all network.

### 2.3.1 Layer A: Representation of the posterior probability
The neuronal architecture of Layer A is based on the original model described in Eq. (8), and the network dynamics correspond to the Bayesian calculation presented in Section 2.1. The largest difference from the original model is that the logarithm of the normalization factor in Eq. (8) is approximated by the feedback input from Layer B (the thick blue arrow in Fig. 1). The output firing rates of the $i$-th excitatory and inhibitory neurons in Layer A are described as follows:

$$A^+_i (t+1) = \alpha w_i I(t) + A^+_i (t) - R(t) - \ln h,$$

$$R(t) = \frac{1}{N^B} \sum_{j=1}^{N^B} b^-_j (t),$$

where $\alpha = 0.3$, and $A^+$ denotes the output firing rate of the excitatory neuron. It should be noted that $A^+$ is equivalent to $V$ in Eq. (8). The right side of Eq. (14) is an approximation of the logarithm.
of the normalization factor, which is the sum of the feedback inputs from the $N_B$ inhibitory neurons in Layer B. As with the original model, the global inhibitory input $R(t)$ for the normalization is given after each update. After normalization, the output firing rate of the neuron is less than 0. Therefore, $\ln h$ is added to $R(t)$ so that the range of firing rates is maintained at a value greater than 0; we initially set $h$ to 0.001. The probability of each excitatory neuron encoding is decoded as $P(\theta_i^j | I(t), ..., I(1)) = h \exp (A_i^+(t))$, where the range of probability neuron encodings is between $h$ and 1.

### 2.3.2 Layer B: Parallel processing for the normalization

As shown in Fig. 1, Layer B consists of $N_B$ neuron groups, and the architecture of each neuron group is a simple winner-take-all (WTA) network. Each of the groups consists of one inhibitory neuron and $N_A$ excitatory neurons. The dynamics of the $i$-th group are described by the following discrete form:

$$b^+_ij(t+1) = \max \left( 0, b^+_ij(t) + \beta \left( a^+_ij(t) + \epsilon_{ij} - b^-_i(t) \right) \right),$$  
$$b^-_i(t+1) = b^-_i(t) + \beta \left( \sum_{j=1}^{N_A} b^+_ij(t) - b^-_i(t) \right),$$

where $\beta$ is the integration rate and $\beta = 0.6$, and $a^+_ij(t) = \alpha w_j I(t) + A^+_ij(t)$. The max function in Eq. (15) outputs the largest value of the elements. It should be noted that $a^+$ is equivalent to $v$ in Eq. (8). $b^+_ij$ and $b^-_i$ are the output firing rates of the $j$-th excitatory and inhibitory neurons in the $i$-th group, respectively. In addition, $\epsilon_{ij}$ represents the background input to the $j$-th excitatory neuron and follows the Gumbel distribution $f(x; -\gamma, 1)$. In each neuron group, the maximum of composite value $(a^+_ij + \epsilon_{ij})$ is sampled in parallel. That is, $b^-_i$ is described as follows when the WTA network reaches a steady state:

$$b^-_i(t+1) = a^+_ik(t) + \epsilon_{ik}, \text{ where } k = \arg \max_j \left\{ a^+_ij(t) + \epsilon_{ij} \right\}.$$  

From Eq. (12), $R$ in Eq. (14) approximates the expectation of the maximum composite value, i.e., the logarithm of the normalization factor, at steady state as follows:

$$R(t) \sim E \left[ \max_i \left\{ a^+_i(t) + \epsilon_i \right\} \right] = \ln \sum_{i=1}^{N_A} \exp \left( a^+_i(t) \right).$$  

### 3. Results

#### 3.1 Orientation discrimination task

Similar to the previous study, we show how the proposed model works as the neural network for Bayesian estimation by demonstrating an orientation discrimination task. In the task, the model network receives orientation stimuli and estimates a stimulus orientation in the presence of significant noise. The sequence of the stimuli was generated by Gabor-filtered images that contain a bright oriented bar and Gaussian noise against a dark background, as shown in Fig. 2(a). The bar orientation was fixed at 0 degrees. The stimulus sequence was received by six excitatory neurons in Layer A. These neurons encode the probability of a bar orientation, which is represented by the discrete states $\theta_i$; the $i$-th neuron encodes the probability of the orientation of $30 \times (i - 1)$ degrees for $i = 1, ..., 6$. The weighted feedforward inputs $w_j I(t)$ in Eq. (13) are replaced with a Gabor filter [6], which provides orientation selectivity. The number of neuron groups in Layer B was initially set to 300.

Owing to the significant noise, it was difficult to immediately determine the presented stimulus orientation from the weighted feedforward inputs $w_j I(t)$ to the neurons in Layer A (Fig. 2(b)). However, as time goes on, the model network advanced the orientation estimation. That is, the output firing rate of the neuron that corresponded to the presented stimulus orientation monotonically increased, and the others decreased (Fig. 2(c)). Accordingly, the probability corresponding to the 0 degree orientation increased and became close to 100% (Fig. 2(d)). In Layer A, the neurons did not connect with each other, but competition between the neurons occurred during the estimation process.
This indirect competition between the neurons in Layer A originated from the inhibitory feedback from Layer B, representing the normalization for encoding the posterior probability in the firing rates. Figure 2(c) shows a time series of the sum of the probabilities encoded by the neurons. When the sum was maintained at 100%, accurate estimates of the posterior probabilities were achieved by the firing rates. Figure 2(f) illustrates the dependence of the sum on the number of groups in Layer B. As shown in the figure, the sum of the probabilities became more stable as the number of neuron groups in Layer B increased. However, the accuracy of the normalization was low during the transient period of estimation regardless of the number of neuron groups. During the transient period, the sum of the encoded probability surpassed the theoretical value. This is because of the dynamics of the WTA network in Layer B.

Fig. 2. Time series of the variables in the proposed model during the orientation discrimination task with significantly noisy visual images. (a) A sample of the noisy visual image. In the image, a horizontal white line exists with Gaussian noise. (b) Feedforward inputs to the neurons in Layer A through a Gabor filter. (c) Outputs of the neurons in Layer A. (d) Probabilities encoded by the neuron in Layer A. (e) Sum of the probabilities. (f) Dependence of the temporal average of the sum of the encoded probabilities on the number of neuron groups in Layer B. The circles denote the temporal averages from 1000 to 1500 time steps, and the bar denotes their standard deviations.

3.2 Behavior of the model to a “random-walk” input
In this section, we demonstrate the behavior of the proposed network model under the conditions in which it is difficult for the network to execute the Bayesian computation. As an example of such a condition, the weighted-feedforward input in Eq. (13) to the neurons in Layer A was given as a fluctuating input, described as $w_i I(t) = 3I_i(t)$. The fluctuating input $I$ was updated as follows:

$$I_i(t + 1) = \max (0, I_i(t) + 0.1 (r_{\text{norm}} - 0.5I_i(t)))$$

where $r_{\text{norm}}$ is the standard normal random value. The number of neurons in Layer A and the number of groups in Layer B were set to 3 and 300, respectively. It was assumed that the neuron in Layer
A encoded the probability about a state of the world, e.g., a motion direction or the position of an external object. Figure 3 shows the time series of the variables in the proposed model.

As shown in Fig. 3, the firing rates of the neurons in Layer A (Fig. 3(b)) and its encoded probability fluctuated (Fig. 3(c) and 3(d)) because the weighted-feedforward inputs to Layer A fluctuated (Fig. 3(a)). In particular, at the time when the Layer A neuron with the largest firing rate was replaced by another, the sum of the probabilities encoded by the neurons surpassed the theoretical value, i.e., normalization failed. This is because the WTA network in Layer B did not necessarily detect the maximum value in each neuron group (Eq. (17)) immediately after the maximum input was replaced. That is, at the switching timing, the maximum value detected by the WTA networks in Layer B could not follow the rapid change in the inputs from the neurons in Layer A well. As a result, a neuron outputted a transient value that was below the theoretical one, and the network failed to calculate the normalization factor exactly. However, in other words, it also played a role in the change-point detection of the external inputs. On the other hand, the network successfully performed Bayesian computation including normalization under conditions in which the input fluctuations were not intensive and the frequency of the turnovers of the inputs was low.

![Fig. 3. Time series of the neural activity in the model to “random-walk” inputs. (a) Feedforward inputs to the neurons in Layer A. (b) Outputs of the neurons in Layer A. (c) Probabilities encoded by the neurons in Layer A. (d) The sum of the probabilities.](image)

### 3.3 The dependency of the model’s performance on the input profile

In this section, we present the dependence of the model performance on the profile of the feedforward input, i.e., the input intensity and bias. The number of neurons in Layer A and the number of groups in Layer B were set to 3 and 300, respectively. The weighted-feedforward input in Eq. (13) to the $i$-th neuron in Layer A was given as a constant input, which is described as
\[ w, I(t) = gr_1, \quad r_i = \begin{cases} 
1/N_A + \delta & i = 1 \\
1/N_A & i = 2 \\
1/N_A - \delta & i = 3
\end{cases}, \] (20)

where \( g, N_A, r, \) and \( \delta \) denote the intensity of the feedforward input, the number of neurons in Layer A, the input rate, and the bias in the inputs, respectively.

Figure 4(a) shows the required number of time steps for estimation from the feedforward inputs regulated by \( g \) and \( \delta \) in Eq. (20). The required number of time steps is an ensemble average of 50 trials. It is assumed that the estimation is completed when the absolute difference between the temporal average of \( h_\text{exp} (A_1(t)) \) for last 50 time steps and its theoretical value, i.e., 1, is less than 0.05.

As the input intensity and bias were smaller, more time steps were required for estimation. Figure 4(b) shows the relations between several values of the input intensity and estimation time. The reaction time of a cognitive task has been known to obey Piéron’s law, which is described by the following equation:

\[ RT = \alpha I^{-b} + c, \] (21)

where \( RT \) and \( I \) are the reaction time and input intensity, respectively [7]. In Fig. 4(b), we regarded the estimation time and input intensity \( g \) as \( RT \) and \( I \) in Piéron’s law and obtained the regression curves of \( f(g) = ag^{-b} + c \). By comparing Fig. 4(b) and (c), the estimation time became shorter as the input intensity increased. However, even though the input intensity changed, there was no difference in the response level (firing rate) of the Layer A neurons after completion of estimation. This means that the network estimated the target state independently of the input intensity.

![Figure 4](image)

Fig. 4. Dependence of the model’s performance on the properties of the feedforward inputs to Layer A. (a) Dependence of the required number of time steps for estimation on the input intensity and bias between inputs. (b) Relation between the ensemble average of the required number of time steps and the input intensity. (b) illustrates the data enclosed by red, green, and purple rectangles in (a). (c) Ensemble averages of the output firing rate \( \langle A_1(t_c) \rangle \) for 50 trials, where \( t_c \) is the time when the estimate is completed.

4. Discussion

As it has been suggested that normalization underlies sensory, especially visual, information processing [8], it is important to understand how normalization is implemented in a neural system. We
proposed a mechanism for the normalization of the posterior probabilities in Bayesian estimation that are encoded by neurons. The mechanism enabled the computation of the logarithm of the sum of the posterior probabilities, as shown in the DC method. The proposed model network successfully performed a demonstration of an orientation discrimination task, in which it was confirmed that the mechanism ensured competition between hidden states.

4.1 Biological plausibility of the proposed model

For the implementation of Bayesian computation including normalization, the following two assumptions are important for the proposed model. First, the model network formed a columnar structure (Fig. 1), and the supporting network, i.e., Layer B, consisted of several WTA networks. The neuronal architecture of the proposed model was based on the physiological knowledge that the cerebral cortex has a columnar structure as well as multiple layers. In the primary visual cortex (V1), neurons that have a similar orientation selectivity form a column [9], and each column consists of six functionally distinct layers: Layers 1–6. In particular, the neurons in Layer 4 receive visual inputs from the retinal circuits through lateral geniculate nuclei (LGN). Layer 4 has excitatory synaptic connections to Layers 2/3, and the excitatory neurons in Layer 4 receive inhibitory feedback input from Layers 2/3 through an indirect pathway via Layer 6 and the inhibitory neurons in Layer 4 [10]. Therefore, Layers A and B of the proposed network model would correspond to Layer 4 and the other layers in V1, such as Layers 2/3 and 6. In addition, Layers 2/3 include microcircuits that consist of lateral connections of pyramidal neurons [11], which would enable implementation of the WTA function [12]. Second, the background synaptic input $\epsilon$ in Eq. (15) was supposed to obey a long-tailed distribution, i.e., the Gumbel distribution, as described in Eq. (10). Equation (12), which is the main point of the proposed model, strictly holds in the case that the background noise obeys the Gumbel distribution; the long-tailed shape of the distribution contributes to the equality of Eq. (12). Therefore, it should be difficult to replace the background synaptic input with a Gaussian noise input. However, the same result could be obtained in the case where the background noise obeys some other long-tailed distribution, e.g., a log-normal distribution. Regarding this point, theoretical and physiological studies show that the excitatory postsynaptic potentials in the neocortex obey a log-normal distribution [13–16]. Indeed, the distributions are different but commonly have long-tailed shapes, suggesting that extremely large inputs might play an important role in the computation.

4.2 Psychophysical plausibility of the proposed model

As shown in Fig. 4, the model’s performance depends on the input profile. In particular, the relation between the time required for estimation and the input intensity $g$ was explained by a power-law relation, as shown in Fig. 4(b). This relation is called Piéron’s law (Eq. (21)) in the field of psychophysics and explains the relation between mean response times and a stimulus intensity [17, 18]. We found that the original model as well could reproduce the psychophysical law. Therefore, it is implied that Piéron’s law should be attributed to the nature of the Bayesian estimation algorithm. Compared with the original model, the proposed model did not necessarily realize a faithful implementation of the Bayesian estimation. However, the proposed model approximated it on a biologically plausible architecture and successfully connects physical stimuli from the external world with perception and cognition.

References

[1] D. Kersten, P. Mamassian, and A. Yuille, “Object perception as Bayesian inference,” Annu. Rev. Psychol., vol. 55, pp. 271–304, 2004.
[2] K.P. Koding and D.M. Wolpert, “Bayesian integration in sensorimotor learning,” Nature, vol. 427, no. 6971, pp. 244–247, 2004.
[3] K. Doya, S. Ishii, A. Pouget, and R.P.N. Rao, Bayesian Brain: Probabilistic Approaches to Neural Coding, MIT Press, Cambridge, Massachusetts, 2007.
[4] R.P.N. Rao, “Bayesian computation in recurrent neural circuits,” Neural Comput., vol. 16, no. 1, pp. 1–38, 2004.
[5] K.E. Train, *Discrete choice methods with simulation*, Cambridge University Press, Cambridge, U.K., 2003.

[6] J.P. Jones and L.A. Palmer, “An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex,” *J. Neurophysiol.*, vol. 58, no. 6, pp. 1233–1258, 1987.

[7] H. Piéron, “Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices,” *Annae Psychol.*, vol. 22, no. 1, pp. 17–96, 1914.

[8] C. Matteo and D.J. Heeger, “Normalization as canonical neural computation,” *Nat. Rev. Neurosci.*, vol. 13, pp. 51–62, 2012.

[9] D.H. Hubel and T.N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[10] G. Stephen, E. Mingolla, and W.D. Ross, “Visual brain and visual perception: How does the cortex do perceptual grouping?,” *Trends Neurosci.*, vol. 20, no. 3, pp. 106–111, 1997.

[11] M. Avermann, C. Tomm, C. Mateo, W. Gerstner, and C.C. Petersen, “Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex,” *J. Neurophysiol.*, vol. 107, pp. 3116–3134, 2012.

[12] R.J. Douglas and K.A. Martin, “Neuronal circuits of the neocortex,” *Annu. Rev. Neurosci.*, vol. 27, pp. 419–451, 2004.

[13] J. Teramae, Y. Tsubo, and T. Fukai, “Optimal spike-based communication in excitable networks with strong-sparse and weak-dense links,” *Sci. Rep.*, vol. 2, 485, 2012.

[14] S. Song, P.J. Sjöström, M. Reigl, S.B. Nelson, and D.B. Chklovskii, “Highly nonrandom features of synaptic connectivity in local cortical circuits,” *PLoS Biol.*, vol. 3, no. 3, e68, 2005.

[15] S. Lefort, C. Tomm, J.C. Floyd Sarria, and C.C.H. Petersen, “The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex,” *Neuron*, vol. 61, pp. 301–316, 2009.

[16] L. Sarid, R. Bruno, B. Sakmann, I. Segev, and D. Feldmeyer, “Modeling a layer 4-to-layer 2/3 module of a single column in rat neocortex: Interweaving in vitro and in vivo experimental observations,” *Proc. Nat. Acad. Sci. USA*, vol. 104, pp. 16353–16358, 2007.

[17] D. Pins and C. Bonnet, “On the relation between stimulus intensity and processing time: Piéron’s law and choice reaction time,” *Percept. Psychophys.*, vol. 58, no. 3, pp. 390–400, 1996.

[18] L.V. Maanen, R.P.P.P. Grasman, B.U. Forstmann, and E.J. Wagenmakers, “Piéron’s law and optimal behavior in perceptual decision-making,” *Front. Neurosci.*, vol. 5, 143, 2011.