White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/176129/

Version: Accepted Version

Article:
Fang, Y, Yu, Z, Liu, JK et al. (2019) A unified neural circuit of causal inference and multisensory integration. Neurocomputing, 358. pp. 355-368. ISSN: 0925-2312

https://doi.org/10.1016/j.neucom.2019.05.067

© 2019, Elsevier. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can’t change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

A unified neural circuit of causal inference and multisensory integration

Ying Fang[a,b], Zhaofei Yu[a,b,c], Jian K. Liu[d], Feng Chen[a,b,e,*]

[a] *Center for Brain-Inspired Computing Research, Department of Automation, Tsinghua University, Beijing 100084, China*
[b] *Beijing Innovation Center for Future Chip,Beijing 100084, China*
[c] *National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing, China*
[d] *Centre for Systems Neuroscience, Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK*
[e] *Beijing Key Laboratory of Security in Big Data Processing and Application, Beijing 100084, China*

Abstract

Causal inference and multisensory integration are two fundamental processes of perception. It is generally believed that there should be one unified neural circuit in the brain to realize these two processes in an optimal way. However, there is no solution yet due to the complicated neural implementation for posterior probability computation. In this study, we propose a unified neural network by solving the complicated posterior probability computation. A unified theoretical framework is presented from the viewpoint of expectation. In addition, a biologically realistic neural circuit is proposed with the combination of importance sampling and probabilistic population coding. Theoretical analyses and simulation results manifest that our proposed neural circuit can implement both causal inference and multisensory integration. Taken together, our framework provides a new perspective of how different perceptual tasks can be performed by the same neural circuit.

Keywords: Multisensory integration, Causal inference, Unified neural circuit, Importance sampling, Probabilistic population codes

*Corresponding author. Tel: (+86)10-62797145
Email address: chenfeng@mail.tsinghua.edu.cn (Feng Chen)

1. Introduction

Perception is fundamental for accurate prediction and efficient action. There are two important parts in perception: causal inference and multisensory integration [1, 2, 3, 4]. For instance, the accurate seeing of localization is crucial for a predator. When the predator receives noisy sensory information from multiple modalities, such as visual and auditory modalities, it should decide whether there is a common cause or not. This process is termed causal inference [4, 5]. When the predator need to localize one object, it will combine sensory signals, i.e. unisensory stimulus, multiple stimuli of the same modality or multiple modalities, into the perception of localization. This process is termed multisensory integration [6, 7]. It has been shown that a number of cognition and perception activities in the brain are processed and inferred in a near-optimal way [8, 9, 10, 11]. As a result, the studies based on Bayesian approach are widely conducted to provide possible neural mechanismes for different issues [12, 5, 13, 14, 15, 16, 17, 18, 19, 20, 21].

In recent years, theoretical ideas about possible neural implementation of causal inference and multisensory integration are proposed, but a separate way.

The core problem of multisensory integration is stimulus estimation by probability. Beck et al. [22] and Ma et al. [23, 24] proposed to use probabilistic population codes to realize neural circuit of probability. The firing rates of a population of neurons correspond to the probability distributions. The linear sum of the input population corresponds to the product of the two posterior distributions. Therefore, they can estimate stimulus by taking the summation of the firing rates of the corresponding neurons in each population. However, they assume that two populations of neurons share the same tuning curve profile, which is the expected value of firing rates, the two populations encode information about the *same* stimulus. In other words, this process is just a single-modal multisensory integration. Peña, Cazettes et al .[25, 26, 27] proposed to use population vector to construct neural networks. They assumed that neurons encode the likelihood probability in their tuning curves and that the density of neurons reflects the prior information. As a result, they can estimate stimulus by the neural implementation of the dot product. However, there should be only one single vector for calculation due to the rule of dot product. Therefore, they computed position just in single sensory condition.

The core problem of causal inference is the calculation for probability of

cause. Ma et al. [28, 29] defined a log posterior ratio to avoid the problematic calculation of exponential function in the posterior probability. Such a simplified calculation can be realized by rational neural activities. As a result, such neural circuits produces optimal decision rule to identify whether there is a common cause. However, when computing stimulus estimate in multisensory integration, their approach can not avoid the calculation of exponential function so that their approach needs extra complexity to deal with exponential function. As a result, the circuit of multisensory integration is different from that of causal inference. Yu et al. [30] constructed a hierarchical neural circuit with importance sampling. This neural circuit encodes both decision and posterior probability, but without estimating stimulus for multisensory integration. Cuppini et al [31] added a downstream layer on the original multisensory neural circuit [32, 33, 34] to perform causal inference. This downstream layer identifies the number of peaks from the multisensory layers to output the number of cause. As a result, these two processes can not be simulated with the same network.

In summary, these previous formalizations of multisensory integration and causal inference are carried out separately. However, the space complexity of one unified neural circuit tend to be lower under the same performance. In addition, the unified circuit will be more flexible as it can perform these two processes independently. Furthermore, it is generally believed that there should be one unified neural circuit in the brain to realize the two processes[35], because the brain tend to make full use of a limited number of neurons in the long-term biological evolution. Here, we present the first unified neural network model of perception that can perform both causal inference and multisensory integration independently and optimally. Our main contributions are threefold.

Firstly, we propose a unified theoretical formalization of these two processes. We prove that both the probability calculation in causal inference and the stimuli estimation in multisensory integration can be expressed as the expectation defined on the posterior distribution. When the random variable in expectation is probability, it represents causal inference. When the random variable in expectation is stimulus, it represents multisensory integration. In this way, adjusting the random variable (i.e. probability or stimulus) in expectation can perform causal inference or multisensory integration, respectively, in the same neural circuit.

Secondly, we propose an efficient neural implementation on this unified theoretical formalization. How to realize the neural implementation of the

complicated exponential posterior distribution is always a challenge. The bottlenecks are mainly due to the inefficient marginalization and the complicated calculation of exponential posterior distribution. First, in order to overcome the above bottlenecks, we use importance sampling to convert the posterior calculation into the easy prior sampling and easy likelihood ratio. Then, we assume that the tuning curve of the neuron is proportional to likelihood and neurons follow some prior distribution in the brain. As a result, we establish the bridge between the neural circuit and the posterior distribution.

Thirdly, our study provides mathematical explanations for the underlying neural underpinnings in the brain computation. Divisive normalization of firing rate in the cortex proposed by neuroscience [36, 37] studies can be linked to the posterior distribution in Bayesian inference. In addition, physiological experimental observations show that synaptic pulse signals in the cerebral cortex change rapidly from one stable pattern to another [38, 39]. In our simulation, when causal inference is switched to multisensory integration, the random variable in expectation is switched from probability to stimulus or vice versa. As a result, the weights of corresponding neural circuit vary rapidly. This simulation results is consistent with the above physiological experimental observations. It further implies that there should be one unified neural circuit in the brain to realize the two processes.

This paper is organized as follows. In Section 2, we present the generative model for causal inference and multisensory integration and introduce the inference algorithm based on importance sampling. In Section 3, based on the generative model in Section2, a unified theoretical framework for two processes is derived and the corresponding neural circuit is designed. Numerical simulations are presented and analyzed in Section 4. Section 5 gives summary and discussion. Theorem derivation is deferred to the Appendix.

2. Bayesian generative model for two processes and importance sampling inference algorithm

To model causal inference and multisensory integration, we consider the following situation. When the observer are exposed to simultaneous auditory and visual stimulus, he should report both the location of stimulus and existence of a common cause. As illustrated in Fig.1, the stimulus can be visual S_1 and auditory S_2) cue. Due to noisy observation, there is a uncertainty in the information conveyed by stimulus, which can be termed as noisy measurement X_1 for visual cue and X_2 for auditory cue. Two processes can be

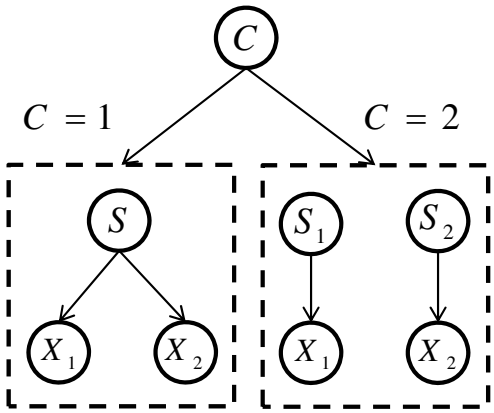

Figure 1: Bayesian generative model of causal inference and multisensory integration.

expressed as:

- casual inference: inferring the existence of a common cause C through noisy measurements X_1 and X_2, where $C = 1$ means the cues have the same cause, and $C = 2$ means the cues have two different causes.
- multisensory integration: inferring the stimuli $(\hat{S}_1, \hat{S}_2)$ that are the optimal estimations from real stimuli $(S_1,\ S_2)$.

When $C = 1$, there is one stimulus S, which is drawn from Gaussian distribution $p(S)$ with mean as 0 and standard deviation as σ_S. This stimulus produces two measurements X_1 and X_2 drawn from Gaussian distributions with the same mean as S but different standard deviation as σ_1 and σ_2. Thus the likelihood probabilities are

$$p\left(X_i|S\right) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\left(X_i - S\right)^2}{2\sigma_i^2}\right)$$

with $i = 1, 2$, respectively.

When C=2, there are two stimuli S_1 and S_2, which are drawn from Gaussian distribution $p(S)$ with mean as 0 and standard deviation as σ_S. Stimuli S_1, S_2 produce two measurements X_1 and X_2 independently drawn from Gaussian distributions with mean as S_1, S_2 and standard deviation as σ_1, σ_2,

respectively. Thus the likelihood is

$$p\left(X_i|S_i\right)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{\left(X_i-S_i\right)^2}{2\sigma_i^2}\right)$$

with $i=1,2$, respectively.

Then, we introduce the inference algorithm based on importance sampling. Given a noisy observation x and true stimulus x^*, the expectation of some function $f(x^*)$ over the posterior distribution $p(x^*|x)$ is often useful. However, it is hard to evaluate the expectation due to inefficient integration and complicated posterior distribution where direct sampling is difficult. A Monte Carlo method, namely importance sampling, can solve such intractable calculation in the elegant way. As Eq.(1) shows, by drawing a large number of samples from the simple prior, importance sampling can approximate the expectation where the posterior distribution is skillfully converted to the ratio of likelihood.

$$\begin{aligned}
E[f(x^*)|x] &= \int f(x^*)p(x^*|x)dx^* \qquad (1)\\
&= \int f(x^*)\frac{p(x^*)p(x|x^*)}{\int p(x^*)p(x|x^*)dx^*}dx^*\\
&\simeq \frac{1}{M}\sum_{i=1}^{M} f(x_i^*)\frac{p(x|x_i^*)}{\int p(x^*)p(x|x^*)dx^*}\\
&\simeq \sum_{i=1}^{M} f(x_i^*)\frac{p(x|x_i^*)}{\sum\limits_{i=1}^{M} p(x|x_i^*)} \qquad x_i^*\sim p(x^*)
\end{aligned}$$

3. One unified theoretical framework for tow processes

Based on the Bayesian generative model in Section2, the aim is to obtain an optimal estimation of the stimulus location and feasible probability calculation of the common cause. Now we propose a unified theoretical framework of two processes and prove that both the probability calculation in causal inference and the stimuli estimation in multisensory integration can be expressed as the expectation defined on the posterior distribution.

Previous studies found that, to perform causal inference with importance sampling, the posterior probability $p(C=1|X_1=x_1,X_2=x_2)$ can be

written as a form of expectation by inducing hidden variables S_1, S_2 as follows [30, 40],

$$\begin{aligned}
\mathrm{p}(\mathrm{C}=1|X_1=x_1, X_2=x_2) =& \mathrm{p}(\mathrm{C}=1|x_1,x_2) \qquad (2)\\
=& \int_{S_1,S_2} \mathrm{p}(\mathrm{C}=1, S_1, S_2|x_1,x_2) dS_1, S_2\\
=& \int_{S_1,S_2} \mathrm{p}(\mathrm{C}=1|S_1,S_2) p(S_1,S_2|x_1,x_2) dS_1, S_2\\
=& E(p(\mathrm{C}=1|S_1,S_2))_{p(S_1,S_2|x_1,x_2)}\\
\simeq& \sum_{S_1^i,S_2^i \sim p(S_1,S_2)} p(\mathrm{C}=1|S_1^i,S_2^i) \frac{p(x_1,x_2|S_1^i,S_2^i)}{\sum_{S_1^i,S_2^i \sim p(S_1,S_2)} p(x_1,x_2|S_1^i,S_2^i)}\\
=& \sum_{S_1^i,S_2^i \sim p(S_1,S_2)} I(S_1^i = S_2^i) \frac{p(x_1,x_2|S_1^i,S_2^i)}{\sum_{S_1^i,S_2^i \sim p(S_1,S_2)} p(x_1,x_2|S_1^i,S_2^i)}.
\end{aligned}$$

Note that in Eq.(2), we abbreviate $X_1 = x_1, X_2 = x_2$ to x_1, x_2 and this will hold in the rest of the paper. $I(S_1^i = S_2^i)$ is an indicator function. When $S_1^i = S_2^i$, it equals to 1. In other cases, it equals to 0.

We proved that causal inference can be inferred in an optimal way by using a sampling-based inference algorithm [30]. Now we show that this algorithm can be used for multisensory integration to estimate S_1.

Without loss of generality, here we estimate S_1. Assuming that the cost of estimation is the mean squared error in Eq.(3), then the optimal estimation is to get the lowest cost under the posterior probability [5]:

$$cost_1 = \int (\hat{S}_1 - S_1)^2 p(S_1|x_1,x_2) dS_1, \qquad (3)$$

where $\hat{S}_1$ is the optimal estimation. Considering the cost is differential, quadratic, and convex, the optimal strategy is to take the derivative with respect to $\hat{S}_1$. Therefore, the optimal estimate is the expectation of stimulus under the posterior probability $p(S_1|x_1,x_2)$ as shown in Eq.(4).

$$\hat{S}_1 = \int S_1 p(S_1|x_1,x_2) dS_1 = E[S_1|x_1,x_2] \simeq \sum_{S_1^i} S_1^i \frac{p(x_1,x_2|S_1^i)}{\sum_{S_1^i} p(x_1,x_2|S_1^i)}. \qquad (4)$$

Comparing the last equation in Eq.(4) with the last equation in Eq.(2), the difference is mainly the likelihood $p(x_1,x_2|S_1^i)$ and $p(x_1,x_2|S_1^i,S_2^i)$. To

reconcile with the theoretical framework of causal inference in Eq. (2), it is necessary to convert the ratio formed by the likelihood $p(x_1, x_2|S_1^i)$ to $p(x_1, x_2|S_1^i, S_2^i)$. By inducing the variable S_2 into the ratio, we get,

$$\begin{aligned}
\frac{p(x_1, x_2|S_1^i)}{\sum\limits_{S_1^i} p(x_1, x_2|S_1^i)} &= \frac{\int p(x_1, x_2, S_2|S_1^i) dS_2}{\sum\limits_{S_1^i} \int p(x_1, x_2, S_2|S_1^i) dS_2} \\
&= \frac{\int p(x_1, x_2|S_2, S_1^i) p(S_2|S_1^i) dS_2}{\sum\limits_{S_1^i} \int p(x_1, x_2|S_2, S_1^i) p(S_2|S_1^i) dS_2} \\
&= \frac{\sum\limits_{S_2^j} p(x_1, x_2|S_1^i, S_2^j)}{\sum\limits_{S_1^i} \sum\limits_{S_2^j} p(x_1, x_2|S_1^i, S_2^j)} \qquad S_1^i \sim p(S_1) \;\; S_2^j \sim p(S_2|S_1) \\
&= \frac{\sum\limits_{S_2^i} p(x_1, x_2|S_1^i, S_2^i)}{\sum\limits_{S_1^i} \sum\limits_{S_2^i} p(x_1, x_2|S_1^i, S_2^i)} \qquad S_1^i, S_2^i \sim p(S_2, S_1)
\end{aligned} \tag{5}$$

In statistics, the operation where first S_1^i is sampled from $p(S_1)$ and then S_2^j is sampled from $p(S_2|S_1)$ is equivalent to the one that S_1^i, S_2^i are sampled from $p(S_1, S_2)$ simultaneously. Thus, Eq.(6) is directly derived. Substitute Eq.(5) in Eq.(4), we have,

$$\begin{aligned}
\hat{S}_1 &= \sum_{S_1^i} S_1^i \frac{p(x_1, x_2|S_1^i)}{\sum\limits_{S_1^i} p(x_1, x_2|S_1^i)} = \sum_{S_1^i} S_1^i \frac{\sum_{S_2^i} p(x_1, x_2|S_1^i, S_2^i)}{\sum\limits_{S_1^i} \sum\limits_{S_2^i} p(x_1, x_2|S_1^i, S_2^i)} \\
&= \sum_i S_1^i \frac{p(x_1, x_2|S_1^i, S_2^i)}{\sum\limits_i p(x_1, x_2|S_1^i, S_2^i)} \qquad S_1^i, S_2^i \sim p(S_1, S_2)
\end{aligned} \tag{6}$$

Furthermore, the accuracy for estimation based on sampling-based inference algorithm in Eq.(6) can be proved by the following theorem. This theorem illustrates that the inference will converge to the optimal estimation when the sample size goes to infinity. The full proof is shown in Appendix A.

Theorem 1. *Let $P(C), P(S_1S_2|C), P(X_1|S_1), P(X_2|S_2)$ are the distributions defined on Bayesian network. $S_1^i, S_2^i \sim p(S_1, S_2)$, then for arbitrary small*

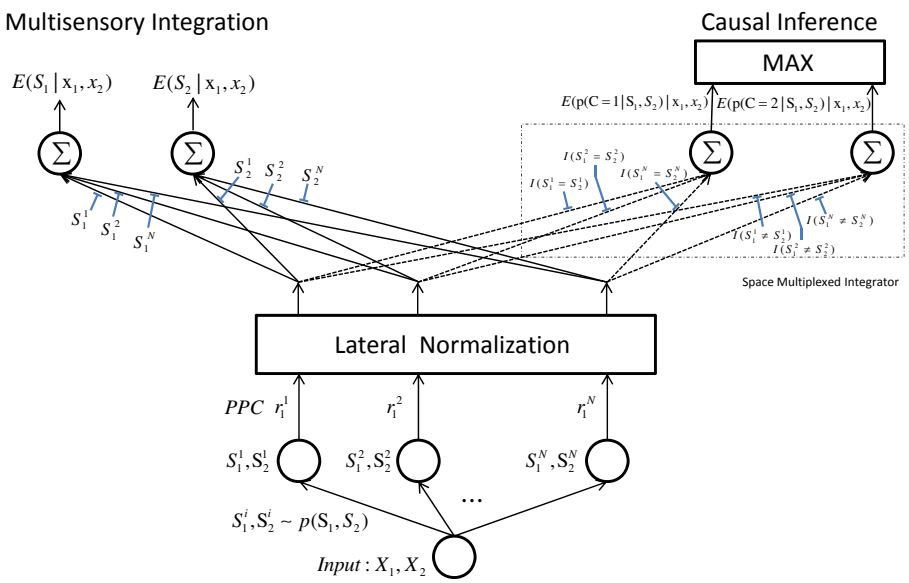

Figure 2: The unified neural circuit of causal inference and multisensory integration.

number ϵ,

$$\lim_{N\to\infty} p(|\sum_{i=1}^{N} S_1^i \frac{p(x_1,x_2|S_1^i,S_2^i)}{\sum_i p(x_1,x_2|S_1^i,S_2^i)} - \int S_1 p(S_1|x_1,x_2)dS_1| < \varepsilon) = 1 \tag{7}$$

Without loss of generality, the estimation of S_2 *also holds.*

As shown in the Eq.(2), the probability of a common cause in causal inference can be expressed as $E(p(\mathrm{C}=1|S_1,S_2))_{p(S_1,S_2|x_1,x_2)}$. As shown in the Eq.(6), the stimuli estimation in multisensory integration can be expressed as $E(S_1)_{p(S_1,S_2|x_1,x_2)}$. It's easy to find that the only difference of these two processes is the random variable in expectation, either probability $p(\mathrm{C}=1|S_1,S_2)$ or stimulus $S_1(or S_2)$. Therefore, we unify the casual inference and multisensory integration into one theoretical framework: the expectation defined on the same posterior distribution. As a result, only adjusting the random variable in expectation can perform causal inference and multisensory integration respectively in the same neural circuit.

4. Neural implementation of the unified framework

A number of studies in psychophysics and physiology suggest that hierarchical Bayesian inference is a reasonable framework to model the brain

activities [41]. Much effort has been devoted to identify a neural substrate to support Bayesian inference model [42, 23, 30, 28, 40]. There are two aspects needed to be considered for neural implementation:

- Encoding: how the stimuli input can generate the activities of neurons;
- Estimation of posterior distribution: how the activities of neurons realize the estimation of posterior probability since we are interested in the expectation over posterior distribution.

For the first problem, a method called probabilistic population coding (PPC) has been introduced to provide neural activities that reflects the information of inputs [23, 42]. In PPC it is assumed that a firing pattern from a population of Poisson spiking neurons encodes a distribution instead of specific values of a variable. The benefit is that the information of underlying quantity is represented by its whole distribution, including variance/uncertainty and other aspects of a distribution. Assuming that Poisson spiking neurons are conditionally independent of each other, this probabilistic encoding is specified as:

$$p(\mathbf{r}|\mathbf{x}) = \prod_i p(r_i|\mathbf{x}) = \prod_i \frac{e^{-f_i(\mathbf{x})} f_i(\mathbf{x})^{r_i}}{r_i!} \tag{8}$$

For neuron i, the tuning curve $f_i(\mathbf{x})$ is the mean firing rate for a range of inputs. The response r_i is the number of spikes in a fixed time interval given the input. In particular, r_i follows the Poisson distribution with the mean firing rate $f_i(\mathbf{x})$ as a parameter. Thus, the neuron activities $r = [r_1, r_2, ..., r_N]$ represent the distribution $p(\mathbf{r}|\mathbf{x})$.

In our work, the characterization of neurons is determined by the prior $p(S_1, S_2)$ under the physiological assumption that the brain follows some prior distributions [43, 44, 45]. Thus, N pairs of neurons $S_1^1 S_2^1, S_1^2 S_2^2, ..., S_1^N S_2^N$ are produced by $p(S_1, S_2)$. Instead of tuning one modal into corresponding activities [40], these neurons $S_1^1 S_2^1, S_1^2 S_2^2, ..., S_1^N S_2^N$ tune multiple modal by choosing the tuning curve $f_i(\mathbf{x})$ as likelihood $p(x_1, x_2|S_1, S_2)$. In this way, these neuron activities are a full firing pattern from multiple stimuli, in some way, to reflect multi-modal uncertainty. Given the stimuli input X_1, X_2, each of neurons $S_1^1 S_2^1, S_1^2 S_2^2, ..., S_1^N S_2^N$ emit spikes $r_i \sim Poisson(c \cdot p(x_1, x_2|S_1, S_2))$. Note that c is some positive constant. According to PPC, the neuron activities $r = [r_1, r_2, ..., r_N]$ encodes the distribution $p(\mathbf{r}|X_1, X_2)$. That is to say, these neuron activities encode the stimuli input X_1, X_2.

For the second problem, it can be divided into two steps, one for calculation of posterior probability and another for neural implementation of such calculation. Calculation of posterior probability is generally difficult. Because it is impractical to sample directly from the posterior distribution with the form of exponential family function. Importance sampling provides a valuable method to approximate the posterior probability. As shown in Eq.(9), it is easy to draw samples (s_1^i, s_2^i) from the prior $p(S1, S2)$ and the posterior probability can be approximated by the ratio of likelihood.

$$\begin{aligned} p(S_1 = s_1^i, S_2 = s_2^i | X_1, X_2) &= \frac{p(X_1, X_2 | s_1^i, s_2^i) p(s_1^i, s_2^i)}{\int p(X_1, X_2 | S_1, S_2) p(S_1, S_2) dS_1, S_2} \\ &\approx \frac{p(X_1, X_2 | s_1^i, s_2^i)}{\sum\limits_i p(X_1, X_2 | s_1^i, s_2^i)} \end{aligned} \tag{9}$$

Then, we explore the neural implementation of posterior probability. Eq.(10) has been proved[40], where R is total firing rate $R = \sum_i r_i$. Note that divisive normalization $E(r_i/R|R = n)$ in Eq.(10) is generally believed to be prevalent in the cortex by neuroscience experiments [36, 37]. Combining the Eq.(10) and Eq.(9), it is obvious that the normalization of firing rate is an unbiased estimator to the posterior probability. In other words, the brain may use divisive normalization to approximate posterior distribution in Bayesian inference.

$$E(r_i/R|R = n) = \frac{p(X_1, X_2 | s_1^i, s_2^i)}{\sum_i p(X_1, X_2 | s_1^i, s_2^i)} \tag{10}$$

Now we will describe the structure and mechanisms implemented in the network.

- Inference process. As shown in the Fig.2, the network processes external inputs X_1, X_2 together in the bottom-up direction. Inputs X_1, X_2 are first fed into a population of neurons $S_1^1 S_2^1, S_1^2 S_2^2, ..., S_1^N S_2^N$. These neurons tune inputs X_1, X_2 into corresponding neuron responses r^1, r^2, $..., r^N$ with tuning curve proportional to $p(X_1, X_2 | S_1^i, S_2^i)$. Then, neurons in lateral inhibition perform divisive normalization, i.e., $E(r_i/R|R = n)$, to these neuron responses. It has been proposed that there is a clear role for lateral inhibition to perform divisive normalization as gain control in visual processing [46, 47, 37, 48]. The normalized responses are

fed into the next layer with synaptic weights W. For multisensory integration, synaptic weights W are $S_1^i(S_2^i)$. The output neurons sum all the inputs from inhibitory neurons with synaptic weights W and response with results $E(S)_{p(S_1,S_2|x_1,x_2)}$. For causal inference, synaptic weights W are switched to $I(S_1^i = S_2^i)(I(S_1^i \neq S_2^i)$ and the output neurons response with results $E(p(\mathrm{C}=1|S_1,S_2))_{p(S_1,S_2|x_1,x_2)}$. In addition, there should be another MAX layer to discriminate whether sensory signals have a common cause. Therefore, the circuit is multiplexed for causal inference and multisensory integration. In other words, causal inference can be performed independently without relying the multisensory integration and vice versa. It is a significant difference between our model with other connectionist models.

- Generation process. In opposite direction to inference, generation process is to generate sampling neurons $S_1^1 S_2^1, S_1^2 S_2^2, ..., S_1^N S_2^N$. Based on the generative model in Section2, we get that sampling neurons follow $S_1, S_2 \sim \mathcal{N}(0, \sigma_p^2)$. Thus, when C=1, sampling neurons $S_1^i = S_2^i$ and they are drawn from the prior $p(S)$, that is $\mathcal{N}(0, \sigma_p^2)$. When C=2, sampling neurons S_1^i and S_2^i are drawn from the prior $p(S)$ respectively.

5. Network simulations

5.1. Verification on the validity of the model

We evaluate the validity of the model from two points: convergence with increasing sampling size and accuracy with different inputs. This part is a basic simulation so the configuration is as follow: the inputs (X_1, X_2) are the two sources of vision and audition; the prior of a common cause is equal to that of two independent causes, that is, $p(C = 1) = p(C = 2) = 0.5$; the reliability of the two stimulus is the same, that is $\sigma_1 = \sigma_2$. Given each input and each size of sampling neurons, we repeat the simulation 10 runs and results are average over 10 runs. When we test convergence, we increase the size of sampling neurons. The 13 different size of sampling neurons is 100 , 200, 500, 1000, 2000, 4000, 6000, 8000, 10000, 12000, 14000, 16000, 18000. Note that the convergence to optimal estimate is quantified by the mean error $\delta = 1/N \cdot \sum_{i=1}^{N} |\hat{S}_1^i - S_1^i|$, where the index i represents the ith input, $\hat{S}_1$ is the optimal estimate according to the formulation given in [28]. $\hat{S}_1$ is

expressed in Eq.(11).

$$\begin{aligned}\hat{S}_1 &= p(C=1|x_1,x_2)\hat{S}_{1,C=1} + p(C=2|x_1,x_2)\hat{S}_{1,C=2} \\ &= \frac{1}{1+e^{-d}}\frac{\frac{x_1}{\sigma_1^2}+\frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}+\frac{1}{\sigma_S^2}} + \frac{1}{1+e^{d}}\frac{\frac{x_1}{\sigma_1^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_S^2}}\end{aligned} \tag{11}$$

The results of testing on the convergence are shown in Fig.3. Fig.3a) is the convergence in multisensory integration and Fig.3b) is the convergence in causal inference. We find that: 1)For multisensory integration in Fig.3a), the mean error of the stimulus estimate decreases exponentially with the sample size increases to 10^3(Note that both figures are semilogarithmic coordinates). In cases of the sample size is larger than 10^4, the mean error approximates to 0 infinitely and becomes stable. It suggests that redundancy of neurons will appear when the sample size increases to a certain amount. In fact, when there are 2000 sampling neurons, the mean error is reasonably small as 0.1 for two-stimulus estimation. Considering there are millions of neurons in the cerebral cortex, our proposed model can approximate the optimal value infinitely. The performance of causal inference in Fig.3b) is similar to that of multisensory integration.

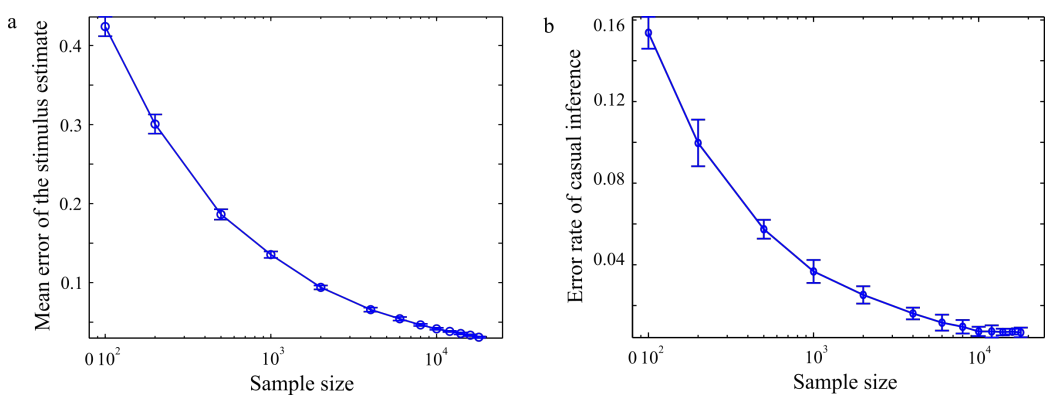

Figure 3: Testing on the convergence. a)Mean error of $|\hat{S}_1 - S_1|$ the stimulus estimate varies with sample size. b)Error rate of causal inference varies with sample size.

When we test accuracy, we produce 1000 inputs (X_1, X_2) with different localization. Some inputs are produced from the common cause and others are not. We compare the trial average of the stimulus estimate S^i with optimal value $\hat{S}^i$ across all inputs for different sample size. The results of

testing on the accuracy are shown in Fig.4. For better display we resort and index the 1000 different inputs according to the optimal value $\hat{S}^1$. It shows that 1)Even if the sample size is 100, 1000 different input estimations all get the satisfactory results. Among them, both the mean error and standard deviation of the 800th input estimation is larger than other input. But when the sample size is 1000, the estimation of 800th input is as accurate as that of others. 2) Given each input, with the sample size increasing, the trial average of the stimulus estimate approximates to the optimal value gradually, and the standard deviation is getting smaller. In a word, when there are 1000 neurons, the approximate result is indistinguishable to the optimal one for each different input.

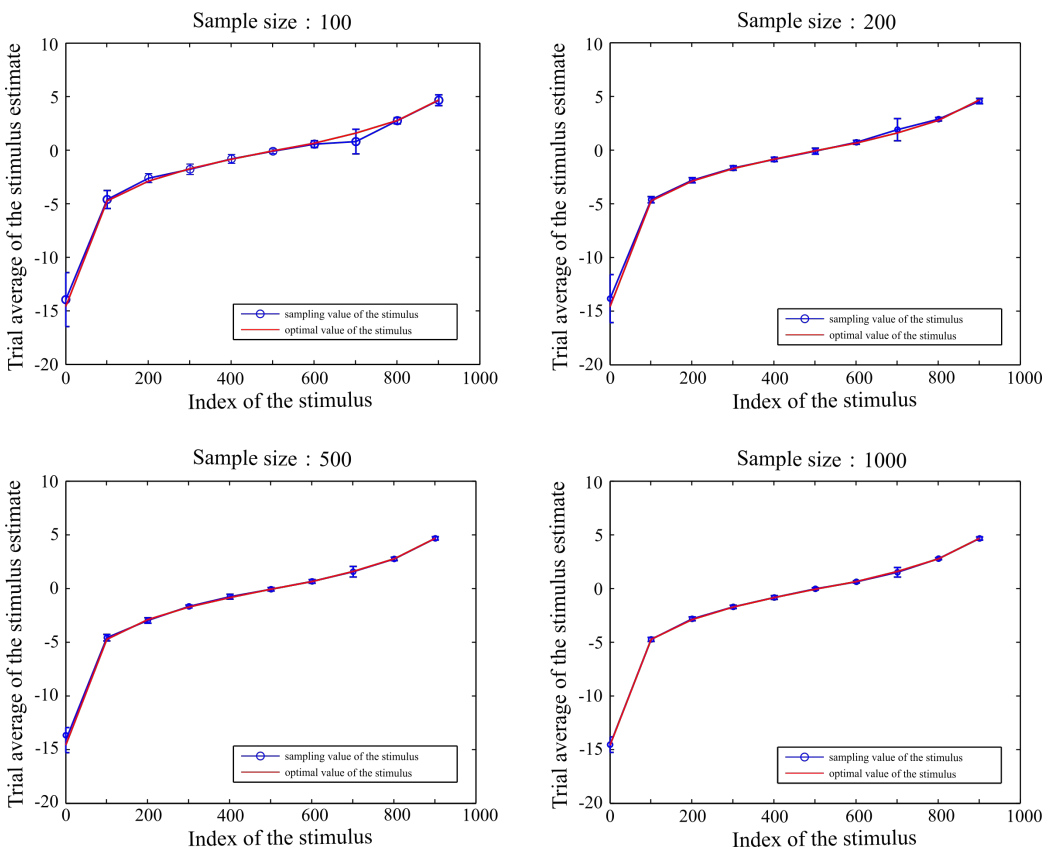

Figure 4: The trial average of the stimulus estimate S comparing with optimal value $\hat{S}$ across 1000 different inputs for different sample size. We repeat the experiment 10 times to get the trial average. For convenience, we resort the stimuli of 1000 different inputs in the increasing order of optimal value $\hat{S}^i$ in Fig.4.

5.2. Verification on the generality of the model

We evaluate the generality of the model from two points: first, it can be applied to unisensory stimulus, multiple stimuli of the same modality or multiple modalities; second, it can be applied to different reliability of sensory information.

In our daily life, the brain receives multiple sources of sensory information, such as visual, auditory, and tactile information. It is necessary to generalize our model to deal with unisensory stimulus, multiple stimuli of the same modality or multiple modalities. As we formalize the causal inference and multisensory integration into the expectation problem defined on conditional posterior, increasing the stimuli only influences the conditional posterior, whose calculation is tractable for our sampling-based method. Thus, it's convenient to implement multi-stimulus task with the same circuit of two-stimulus task. Similar to Eq.(4)-Eq.(6), multiple stimuli estimation is formulated as follows.

$$\begin{aligned}
\hat{S}_k &= E[S_k|x_1, x_2, ..., x_n] \qquad (12)\\
&\simeq \sum_{S_k^i} S_k^i \frac{p(x_1, x_2, ..., x_n|S_k^i)}{\sum\limits_{S_k^i} p(x_1, x_2, ..., x_n|S_k^i)}\\
&= \sum_{S_k^i} S_k^i \frac{\sum\limits_{\{S_1^i, S_2^i, ..., S_n^i\}/S_k^i} p(x_1, x_2, ..., x_n|S_1^i, S_2^i, ..., S_n^i)}{\sum\limits_{S_1^i, S_2^i, ..., S_n^i} p(x_1, x_2, ..., x_n|S_1^i, S_2^i, ..., S_n^i)}\\
&= \sum_i S_k^i \frac{p(x_1, x_2, ..., x_n|S_1^i, S_2^i, ..., S_n^i)}{\sum\limits_i p(x_1, x_2, ..., x_n|S_1^i, S_2^i, ..., S_n^i)} \quad S_1^i, S_2^i, ..., S_n^i \sim p(S_1, S_2, ..., S_n)
\end{aligned}$$

Comparing Eq.6 and Eq.12, it's obvious that we can realize the generalization just by converting several properties of two stimuli to that of multiple stimuli. These properties include: 1), sampling prior $p(S_1, S_2, ..., S_n)$; 2), the tuning curve proportional to $p(x_1, x_2, ..., x_n|S_1^i, S_2^i, ..., S_n^i)$; and 3), the synaptic weights $I(S_1^i = S_2^i = ... = S_n^i)$ $(I(S_1^i \neq S_2^i \neq ... \neq S_n^i))$. Similarly, this model can also be used for one sensory stimulus according to Eq.1. For this, one can simply replace the likelihood $p(X_1, X_2|S_1, S_2)$ by $p(X|S)$, then the sampling neurons become $S^1, S^2, ..., S^N$ from the prior $p(S)$.

We test the spatial accuracy of the model in the unisensory and multisensory conditions where the auditory and visual stimuli are presented in the

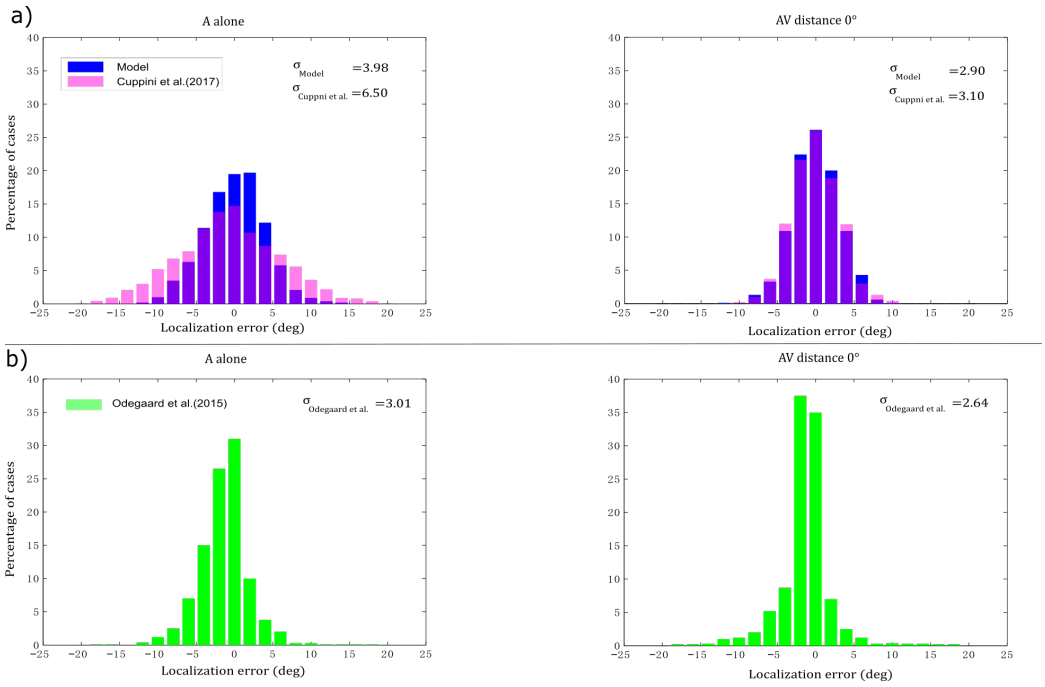

Figure 5: Spatial distribution of the auditory localization error. a) comparison between the our model and Cuppini's model in the unisensory (auditory input alone:left panel) vs. multisensory case (stimuli coincident in space:right panel) b) comparison in the behavioral experiments of Odegaard et al modified from [31]. The inserted text is the corresponding standard deviations.

same position. In addition, we compare the results with those reported by Cuppini et al.[31] and behavioral data by Odegaard et al. [49]. Here, the size of sampling neurons is 1000 and the reliability configuration is the same as Cuppini et al. :$\sigma_{vision} > \sigma_{Auditory}$ as visual stimuli is more reliable. Fig.5 shows that localization error is biased towards the center when two stimuli are coincident. The SD of the auditory localization error in multisensory conditions is significantly improved (which falls from 3.98° in unisensory condition to 2.90° in multisensory conditions). Such a result is in line with the fact that the localization of multisensory integration is more accurate than unisensory estimation. The distribution of the localization of our model is comparable to the behavioral findings of Odegaard et al. (The work by Cuppini et al. and Odegaard et al. will be introduced in Section 5.3.)

We simulated the cases with three stimuli and ten stimuli respectively to test the robustness of our model. The inputs are multiple stimuli that

contain both a single modality and multiple modalities. We adopt the same configurations and the criteria of convergence and accuracy as section 5.1. The results of testing on the convergence are shown in Fig.6. Fig.6a) is the convergence in multisensory integration and Fig.6b) is the convergence in causal inference. It shows that: 1)Three stimuli task and ten stimuli task show the similar behaviors of performance for our model.When the number of sampling neurons increases, the error rate can be arbitrarily small in causal inference, and the approximation manifestly converges to the optimal value in multisensory integration. 2) The results of ten stimuli is not as good as that of three stimuli. It is in line with human behavior that it is difficult to identify too many stimuli. But, results of ten stimuli becomes desirable with adequate sample neurons. The results of testing on the accuracy with ten stimuli are shown in Fig.7. It shows that 1)Given each different input, with sample size increasing, the trial average of the stimulus estimate approximates to the optimal value gradually, and the variance is getting smaller. 2) when the sample size is 100 in Fig.7 a), half of the inputs has a relatively large estimation error, such as, the input index is 0,100,500,600. But the approximation becomes indistinguishable to the optimal one with 1000 sampling neurons in Fig.7 d).

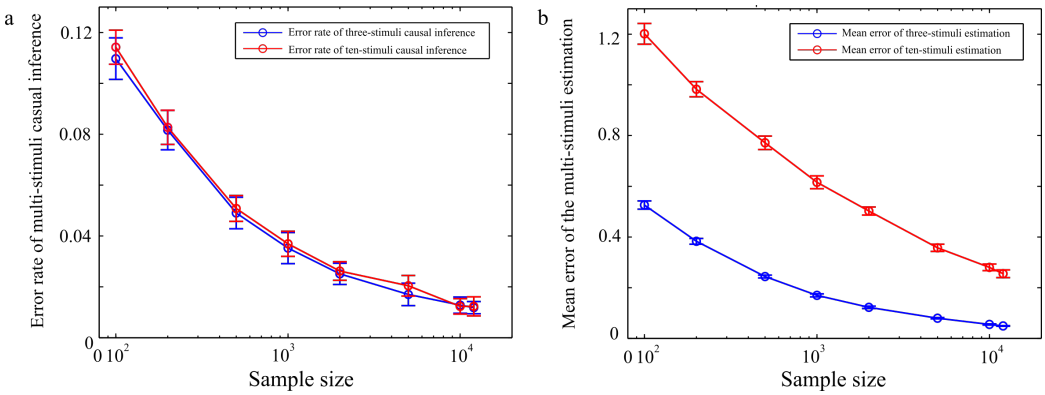

Figure 6: Testing on the convergence in multiple stimuli task. a)Mean error $|\hat{S}_1 - S_1|$ of the stimulus estimate varies with sample size. b)Error rate of causal inference varies with sample size.

Then, we test the generality of the model with different reliability. We change these three parameters of sensory information in a wide range as σ_S, σ_1 and σ_2 varying from 1 to 8 separately. Fig.8 shows the result of multisensory integration: 1) The color graphics are asymmetrical. For example,

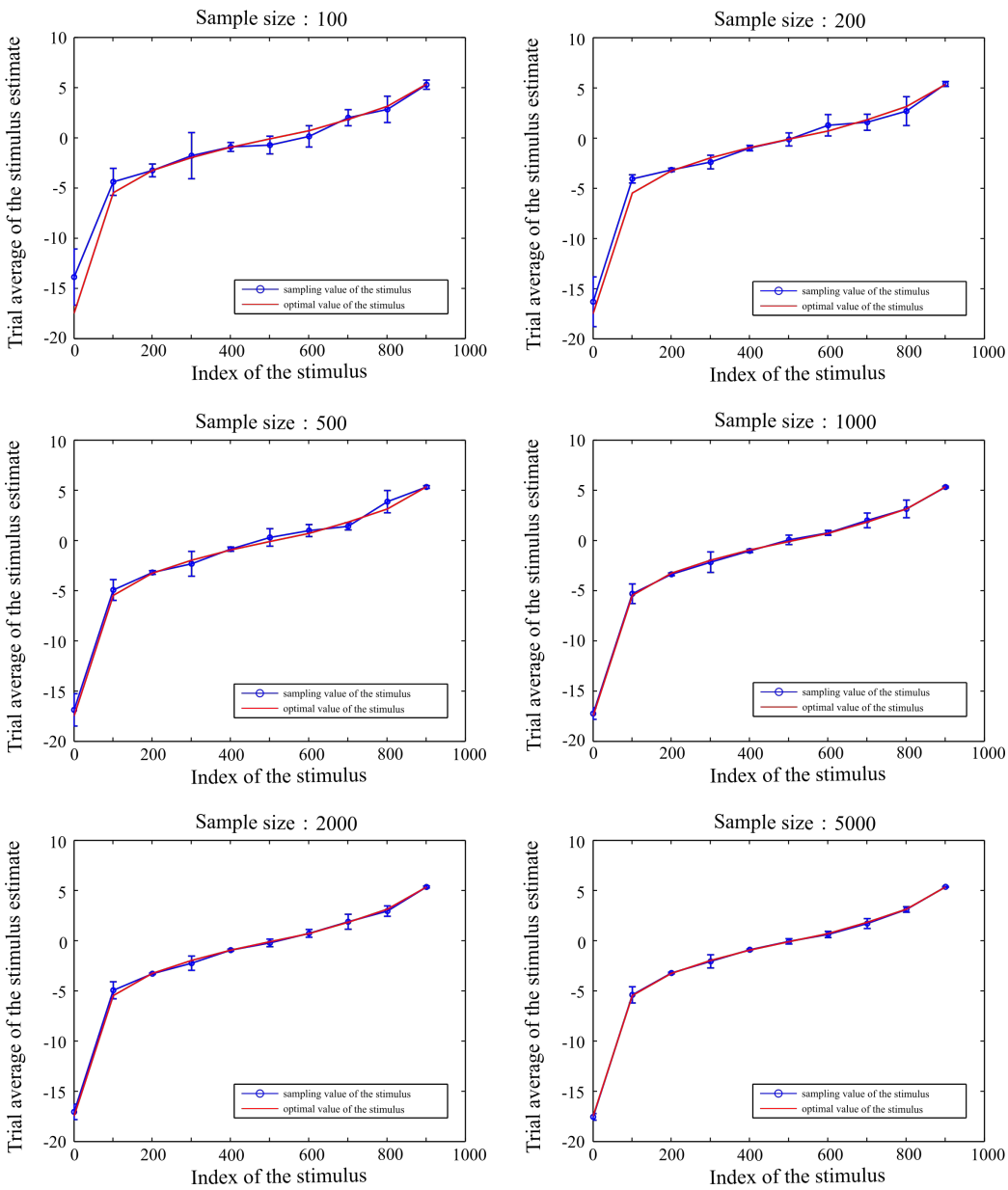

Figure 7: The trial average of the stimulus estimate S comparing with optimal value $\hat{S}$ across the 1000 different inputs for different sample size.

color of most block in the bar $(\sigma_1 = 1, \sigma_2 = 1:8)$ is close to blue while color in the bar $(\sigma_2 = 1, \sigma_1 = 1:8)$ is close to red. Because the estimation of S_1

will be more affected by the reliability of S_1 intuitively. 2) We divide panel a) into four parts. Part in the bottom right corner represents $\sigma_1 < \sigma_2$ and color in this part is close to blue. Part in the upper left corner represents $\sigma_2 < \sigma_1$ and color in this part is close to red. Because intuitively, it is easier to estimate S_1 when S_1 is more reliable than S_2 and estimation of S_1 will be badly affected when S_2 is more reliable. This performance also appears in other panels b), c), d). Fig.9 shows the result of causal inference: 1) The color graphics are symmetrical. Because whether $\sigma_1 < \sigma_2$ or $\sigma_2 < \sigma_1$, there will be no effect on the discrimination. 2) Color in the bottom left part is close to blue. Because when both stimuli are more reliable, it is easier to discriminate multiple sensory signals. 3) When σ_S gets larger, the error rate of causal inference is smaller. Because it is easier to discriminate when the difference between two stimuli is larger.In general, the error rate is less than 0.1 and the mean error is less than 0.3 for the most parameters, which imply that our model has a good performance for different reliability.

5.3. Verification on the applicability of the model

Applicability of our model is reflected in not only performing causal inference and multisensory integration, but also identify other neural mechanism such as spatial ventriloquism [50, 51]. Spatial ventriloquism is a special illusory phenomenon when the brain deals with causal inference and multisensory integration. One of the performances is that when disparity between sensory signals is increasing, the more reliable modality strongly influences the other; when the disparity increases to a certain extent, multisensory integration breaks down. We simulated different experiments compared with a existing connectionist model[31] and behavior data[49]. Here we briefly introduce the above two works.

Odegaard et al.[49] tested the biases in visual and auditory localization in the behavioral experiments. 384 observers answer the position and magnitude in both visual and auditory modalities when stimuli are presented alone or combined with different positions. Data revealed that the visual modality is more reliable than auditory one and distance greater than 5° suggests independent sources.

Cuppini et al.[31] provided a biologically inspired neurocomputational model for audiovisual integration and causal inference. This model consists of three layers: two encode auditory and visual stimuli separately and are reciprocally connected via cross-modal synapses. Then, two layers are connected to the downstream layer. The first two layers realize the multisensory

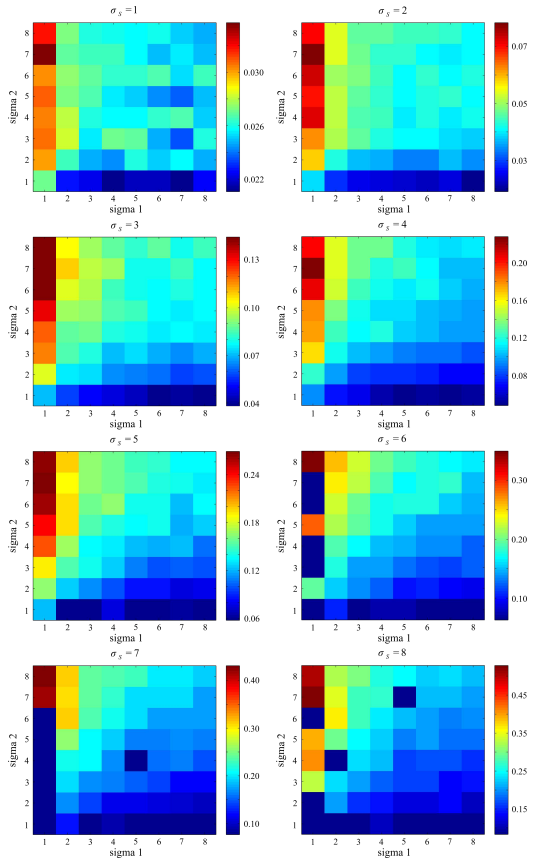

Figure 8: Estimation error $|\hat{S}_1 - S_1|$ for different reliability with sample size being 1000. and the model is fed into 1000 different inputs with the sample size fixed as 1000.

integration while the downstream layer realizes the causal inference. Among others, the network can account for the ventriloquism effect. Cuppni et al simulated different behavioral experiments of sensory detection tasks and compared with corresponding behavior data. And results are in line with human behavior.

In this simulation, we explore the pattern of sensory bias, the report of unity and the distribution of the stimuli localization as a function of the distance between the auditory and visual stimuli. We always fix the position of the visual stimulus and shift the position of the auditory of the auditory stimulus. The distance between two sensory signals changes from 0° to 20°. And we adopt the same configurations as Cuppini et al. Generally, when the stimuli are coincident in space, it is more likely that two sources are

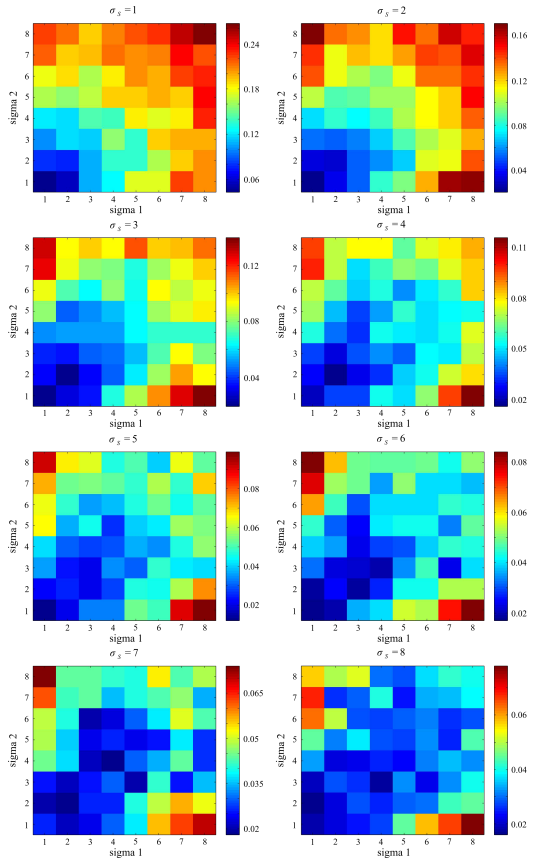

Figure 9: Error rate of casual inference for different reliability with sample size being 1000. and the model is fed into 1000 different inputs with the sample size fixed as 1000.

originated from the common cause on a priori. When the two stimuli are far away, it is more likely that two sources are originated from the different cause on a priori. Therefore, the prior of the common cause decreases from 0.7 to 0.3 with the auditory-visual distance increasing from 0° to 20°. The size of sampling neurons is 1000. Results were evaluated as the mean response over 1000 runs of the same task.

Fig.11 a) shows the report of the unity with the audiovisual spatial disparity increasing. It shows the frequency of one common cause identified by the network. When audiovisual disparity is smaller than 5°, the network judges the two stimuli being one common cause in more than 75% of the simulations. When the disparity is increasing, the percentage of identification of a common cause decreased linearly with the distance. When the audiovisual

t]

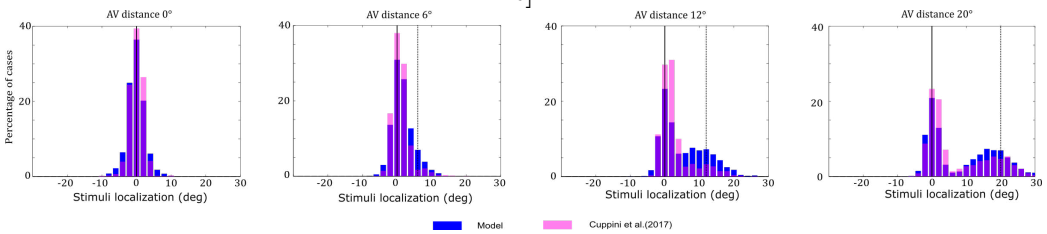

Figure 10: Distribution of the localization of two-modal stimuli at different AV distances simulated with our model and Cuppini's model[31]. The visual stimulus is fixed at 0° (continuous vertical line) and the auditory stimulus is shifted from 0° to 20° (dashed vertical line)

distance is 20°, the network judges the two stimuli being one common cause in almost 30% of the simulations. These results are consistent with human behavior: when the distance between inputs is increasing, human are more likely to identify the existence of different causes.

Fig.11 b) shows the bias of the perceived auditory position. It is computed as the spatial disparity between the real position and evaluated position, divided by the distance between the real auditory and the real visual stimuli. When audiovisual disparity is smaller than 8°, the bias remains stable. While the audiovisual disparity is larger than 8°, the bias decreases linearly.

Fig.10 shows the distribution of the localization of two-modal stimuli at different AV distances. The fixed continuous vertical line is visual stimulus and the shifted dashed vertical line is auditory stimulus in each panel. When the audiovisual disparity is smaller than 10°, the localization is mainly distributed near the position of visual stimulus and the more reliable visual modality dominates in the spatial domain; when the audiovisual disparity is larger than 10°, a bimodal distribution becomes evident: the first peak centers at the real position of visual stimulus and the second peak centers at the real position of auditory stimulus. The results are in accordance with human behavior: when the audiovisual disparity is within a certain range, the more reliable stimulus attract the other; when the audiovisual disparity increases to a certain amount, multisensory integration break down and estimation of these two stimuli does not affect each other.

In the following, we analyze the space complexity between our model and work of Cuppini et al. Without loss of generality, we assume there are two stimuli of vision and audition receptively, and the number of neurons in one

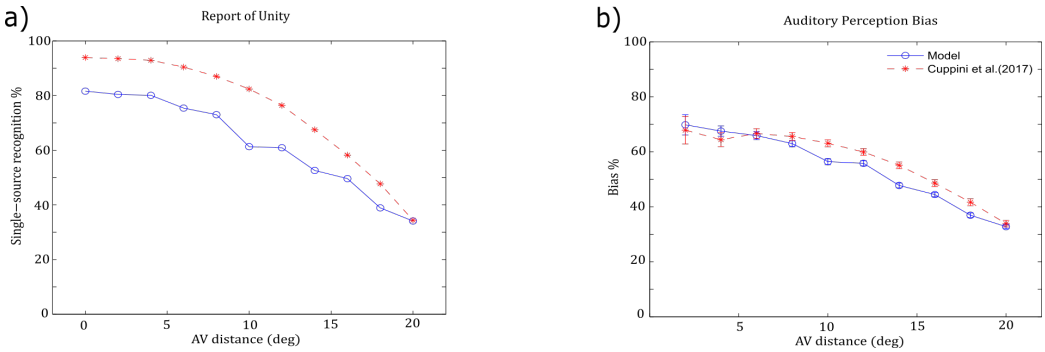

Figure 11: Report of unity (Fig a)) and Auditory Bias (Fig b)) at different AV distances simulated with our model and Cuppini's model[31].

modality is n. In our model, there are two populations of the visual and auditory sampling neurons and the number of output neurons is 4 (2 for causal inference and 2 for multisensory integration). So the total number of neurons is $2n$ (Note that we ignore the output neurons because the number of output neurons is far less than that of sampling neurons). The weight between sampling neurons and output neurons is $4n$ and the weight in the lateral inhibition is n. So the total size of weight is $5n$. The total size of space in our model is $7n$. In the model of Cuppini et al., there are three populations of neurons (two layer for vision and audition respectively and one downstream layer for causal inference). So the total number of neurons is $3n$. The weight between sensory layers and downstream layer is $2n^2$. The weight in cross-modal part is n. The weight in the lateral inhibition is $3n^2$. So the total size of weight is $11n^2+n$. The total size of space in Cuppini et al. is $11n^2+4n$. The table 1 shows the space complexity for better comparison. It shows that the space complexity in our model and Cuppni et al. is $O(n)$ and $O(n^2)$ respectively. Therefore, our model is significantly simple in the case of the same experiment results compared with Cuppini et al.

6. Conclusion

In this paper, we unify the formalization of causal inference and multisensory integration as the expectation defined on the same posterior distribution. Then, based on probabilistic population codes and importance sampling, we provide a reasonable neural circuit to realize these two processes optimally and independently. To our knowledge, this is the first unified neural cir-

Table 1: Comparison of space complexity with our model and Cuppini's model.

	Our model	Cuppini et al.
the size of neuron	$2n$	$3n$
the size of weight	$5n$	$11n^2 + n$
Total size	$7n$	$11n^2 + 4n$
Space complexity	$O(n)$	$O(n^2)$

cuit for these two processes. For mathematical rigor, theoretical analysis and simulation results show that the approximation error can be arbitrarily small when the sample size goes to infinity. Our simulations also show that our circuit can get the satisfactory results with different reliability, unisensory stimulus, multiple stimuli of the same modality or multiple modalities. In addition, spatial ventriloquism can also be appeared in our circuit on the verification of applicability.

Compared with other connectionist biological models, the advantages of our model are two folds. From computational perspective, causal inference can be performed in our model independently without relying on the multisensory integration and vice versa. It suggests that our model will be more flexible. In addition, the space complexity of our unified neural circuit is lower under the same performance. It is consistent with long-term biological evolution in which the brain tend to make full use of a limited number of neurons. From physiological perspective, our unified circuit interprets some underlying neural mechanisms mathematically. Divisive normalization of firing rate in the cortex [36, 37] can be linked to the posterior distribution in Bayesian inference. In addition, rapid weight adjustment observed in the cortex can be linked to the adjustment of variable in expectation when switching between different processes [38, 39].

In this study, the proposed neural circuit only focuses on how to implement two processes, in which synaptic weights are given by a explicit form. A clear question is how synaptic weights are updated to learn the new distribution. Indeed, it is a significant challenge for understanding how the brain adapts itself to the ever-changing world. Based on synaptic sampling, Kappel et al [52] shed new light on the learning process in local neuronal circuit. The essence of synaptic sampling is stochastic gradient descent. The differential term is demonstrated theoretically as optimal STDP (Spike-timing dependent plasticity) learning rule. Thus, the stochastic dynamics of network

parameters enable their networks to learn the new distributions. The benefits of sampling-based computation for spiking neurons and synapses have been recently demonstrated in a series of studies [52, 53, 54, 55]. Future extension of the current model is to address learning with sampling computation of neurons and synapses.

Proof of Theorem 1

In order to prove Theorem 1, we present the Lemma 1 which has been proved in [30].

Lemma 1. Supposing that random variables $X^1, X^2, ..., X^n$ are pairwise independent and $X^i \sim P(X)$. Similarly, $Y^1, Y^2, ..., Y^n$ are pairwise independent and $Y^j \sim P(Y)$. Besides, $E(X) = \mu_1$, $E(Y) = \mu_2$, $\mu_1, \mu_2 \neq 0$, $Var(X) = \sigma_1^2$ and $Var(Y) = \sigma_2^2$. Then for arbitrary small number ε, we can conclude that $P\left(\left|\frac{\sum_{i=1}^{N} X^i}{\sum_{j=1}^{N} Y^j} - \frac{\mu_1}{\mu_2}\right| < \varepsilon\right) > 1 - \frac{16\sigma_1^2}{N\mu_2^2\varepsilon^2} - \frac{16\mu_1^2\sigma_2^2}{N\mu_2^4\varepsilon^2}$.

Theorem 1. Let $P(C), P(S_1S_2|C), P(X_1|S_1), P(X_2|S_2)$ are the distributions defined on the Bayesian network.$S_1^i, S_2^j \sim p(S_1, S_2)$, then for arbitrary small number ϵ,

$$\lim_{N\to\infty} p(|\sum_{i=1}^{N} S_1^i \frac{p(x_1,x_2|S_1^i,S_2^i)}{\sum_i p(x_1,x_2|S_1^i,S_2^i)} - \int S_1 p(S_1|x_1,x_2)dS_1| < \varepsilon) = 1$$

Without loss of generality, the estimation of S_2 also holds for above theorem.

Proof: Supposing that $f_1(x_1,x_2) = \frac{\sum_{i=1}^{N} S_1^i P\left(x_1,x_2|S_1^i,S_2^i\right)}{\sum_{i=1}^{N} P\left(x_1,x_2|S_1^i,S_2^i\right)}$, $f_2(x_1,x_2) = \frac{\sum_{i=1}^{N} S_1^i P\left(x_1,x_2|S_1^i,S_2^i\right)}{\sum_{j=1}^{N} P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)}$, **where** $S_1^i, S_2^i \sim P(S_1,S_2)$ **and** $\tilde{S}_1^j, \tilde{S}_2^j \sim P(S_1,S_2)$**, then**

$$\begin{aligned}
&E\left(\frac{1}{N}\sum_{i=1}^{N} P\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&= \frac{1}{N}\sum_{i=1}^{N} E\left(P\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&= \frac{1}{N}\sum_{i=1}^{N} \int_{S_1^i,S_2^i} P\left(x_1,x_2|S_1^i,S_2^i\right) P\left(S_1^i,S_2^i\right) dS_1^i,S_2^i\\
&= P(x_1,x_2)\\
&= \mu_1,
\end{aligned}$$

Similar to the above derivation, we obtain,

$$E\left(\frac{1}{N}\sum_{j=1}^{N} P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)\right) = P(x_1,x_2) = \mu_1,$$

$$\begin{aligned}
&E\left(\frac{1}{N}\sum_{i=1}^{N} S_1^i P\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&= \int_{S_1,S_2} S_1 p(X_1,X_2|S_1,S_2)p(S_1,S_2)dS_1dS_2\\
&= \int S_1 p(X_1,X_2,S_1)dS_1\\
&= \mu_2,
\end{aligned}$$

It is reasonable to assume that $S_1, S_2 \in [-L, L]$ **and thus expectation** $E\left(\frac{1}{N}\sum_{i=1}^{N} S_1^i P\left(x_1,x_2|S_1^i,S_2^i\right)\right) = \int_{-L}^{L} S_1 p(X_1,X_2,S_1)dS_1$ **is bounded by** D**.**

$$\begin{aligned}
&Var\left(\frac{1}{N}\sum_{i=1}^{N} P\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var\left(P\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&= \frac{1}{N^2}\sum_{i=1}^{N} \left(E\left(P\left(x_1,x_2|S_1^i,S_2^i\right)^2\right) - E\left(P\left(x_1,x_2|S_1^i,S_2^i\right)\right)^2\right)\\
&= \frac{1}{N}\int_{S_1,S_2} P(x_1,x_2|S_1,S_2)^2 P(S_1,S_2)\, dS_1,S_2 - P(x_1,x_2)^2\\
&= \sigma_1^2,
\end{aligned}$$

Likewise, we obtain,

$$Var\left(\frac{1}{N}\sum_{j=1}^{N}P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)\right)=\sigma_1^2,$$

$$\begin{aligned}
&Var\left(\frac{1}{N}\sum_{i=1}^{N}S_1^iP\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&=\frac{1}{N}Var\left(S_1^iP\left(x_1,x_2|S_1^i,S_2^i\right)\right)\\
&=\frac{1}{N}\int_{S_1,S_2}S_1^iP(x_1,x_2|S_1,S_2)^2P\left(S_1,S_2\right)dS_1,S_2-\int S_1p(X_1,X_2,S_1)dS_1{}^2\\
&=\sigma_2^2,
\end{aligned}$$

Note that $S_1,S_2\in[-L,L]$ and thus variance is bounded. [30] has proved that for arbitrary small number ε,

$$P\left(\left|\frac{1}{\frac{1}{N}\sum_{j=1}^{N}P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)}-\frac{1}{\frac{1}{N}\sum_{i=1}^{N}P\left(x_1,x_2|S_1^i,S_2^i\right)}\right|<2\varepsilon\right)\geq\left(1-\frac{\sigma_1^2}{N\mu_1^2\varepsilon^2}\right)^2.$$

We also can get that,

$$\left|f_1\left(x_1,x_2\right)-f_2\left(x_1,x_2\right)\right|=\frac{1}{N}\sum_{i=1}^{N}S_1^iP\left(x_1,x_2|S_1^i,S_2^i\right)\left|\frac{1}{\frac{1}{N}\sum_{j=1}^{N}P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)}-\frac{1}{\frac{1}{N}\sum_{i=1}^{N}P\left(x_1,x_2|S_1^i,S_2^i\right)}\right|$$

$$\leq D\left|\frac{1}{\frac{1}{N}\sum_{j=1}^{N}P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)}-\frac{1}{\frac{1}{N}\sum_{i=1}^{N}P\left(x_1,x_2|S_1^i,S_2^i\right)}\right|$$

We can get for arbitrary small number ε,

$$P\left(\left|f_1\left(x_1,x_2\right)-f_2\left(x_1,x_2\right)\right|<\varepsilon\right)\geq\left(1-\frac{4\sigma_1^2D^2}{NP{\mu_1}^2\varepsilon^2}\right)^2.$$

By virtue of the fact that $\frac{E\left(\sum_{i=1}^{N}S_1^iP\left(x_1,x_2|S_1^i,S_2^i\right)\right)}{E\left(\sum_{j=1}^{N}P\left(x_1,x_2|\tilde{S}_1^j,\tilde{S}_2^j\right)\right)}=\frac{\int S_1p(X_1,X_2,S_1)dS_1}{p(X_1,X_2)}=\int S_1p(S_1|x_1,x_2)dS_1$**, it is easy to use Lemma 1 to show that for arbitrary small number ε,**

$$P\left(\left|f_2\left(x_1,x_2\right)-\int S_1 p(S_1|x_1,x_2)dS_1\right|<\varepsilon\right)>1-\frac{16\sigma_2^2}{N\mu_1^2\varepsilon^2}-\frac{16\mu_2^2\sigma_1^2}{N\mu_1^4\varepsilon^2},$$

We also know that,

$$P\left(\left|f_1\left(x_1,x_2\right)-f_2\left(x_1,x_2\right)\right|<\varepsilon\right)\geq\left(1-\frac{4\sigma_1^2D^2}{NP{\mu_1}^2\varepsilon^2}\right)^2.$$

By application of $P\left(A\cap B\right)\geq P\left(A\right)+P\left(B\right)-1$, for arbitrary small number ε we have,

$$\begin{aligned}&P\left(\left|f_1\left(x_1,x_2\right)-\int S_1 p(S_1|x_1,x_2)dS_1\right|<2\varepsilon\right)\\&\geq\left(1-\frac{4\sigma_1^2D^2}{N\mu_1^2\varepsilon^2}\right)^2-\frac{16\sigma_2^2}{N\mu_1^2\varepsilon^2}-\frac{16\mu_2^2\sigma_1^2}{N\mu_1^4\varepsilon^2}.\end{aligned}$$

When N goes to infinite, it straightforward to conclude that for arbitrary small number ε,

$$\lim_{N\to\infty}p(|\sum_{i=1}^{N}S_1^i\frac{p(x_1,x_2|S_1^i,S_2^i)}{\sum_i p(x_1,x_2|S_1^i,S_2^i)}-\int S_1 p(S_1|x_1,x_2)dS_1|<\varepsilon)=1$$

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 61671266, 61327902, in part by the Research Project of Tsinghua University under Grant 20161080084, and in part by National High-tech Research and Development Plan under Grant 2015AA042306, in part by the National Natural Science Foundation of China under Grant 61806011, in part by National Postdoctoral Program for Innovative Talents under Grant BX20180005, and in part by China Postdoctoral Science Foundation under Grant 2018M630036, and in part by International Talent Exchange Program of Beijing Municipal Commission of Science and Technology under Grant Z181100001018026, and in part by the Royal Society Newton Advanced Fellowship under Grant NAF-R1-191082.

References

[1] M. O. Ernst, H. H. Bülthoff, Merging the senses into a robust percept, Trends in Cognitive Sciences 8 (4) (2004) 162–169.

[2] L. Shams, A. R. Seitz, Benefits of multisensory learning, Trends in Cognitive Sciences 12 (11) (2008) 411 – 417.

[3] J. K. Bizley, G. P. Jones, S. M. Town, Where are multisensory signals combined for perceptual decision-making?, Current Opinion in Neurobiology 40 (2016) 31 – 37.

[4] L. Shams, U. R. Beierholm, Causal inference in perception, Trends in Cognitive Sciences 14 (9) (2010) 425 – 432.

[5] K. P. Körding, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, L. Shams, Causal inference in multisensory perception, PLoS One 2 (9) (2007) e943.

[6] C. Chandrasekaran, Computational principles and models of multisensory integration, Current Opinion in Neurobiology 43 (2017) 25 – 34.

[7] M. Ursino, C. Cuppini, E. Magosso, Neurocomputational approaches to modelling multisensory integration in the brain: A review, Neural Networks 60 (2014) 141 – 165.

[8] P. W. Battaglia, R. A. Jacobs, R. N. Aslin, Bayesian integration of visual and auditory signals for spatial localization, Journal of the Optical Society of America A 20 (7) (2003) 1391–1397.

[9] M. O. Ernst, M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature 415 (6870) (2002) 429–33.

[10] D. M. Merfeld, L. Zupan, R. J. Peterka, Humans use internal models to estimate gravity and linear acceleration, Nature 398 (6728) (1999) 615.

[11] D. M. Wolpert, Z. Ghahramani, M. I. Jordan, An internal model for sensorimotor integration, Science 269 (5232) (1995) 1880.

[12] D. Alais, D. Burr, The ventriloquist effect results from near-optimal bimodal integration, Current Biology 14 (3) (2004) 257–262.

[13] Y. Sato, T. Toyoizumi, K. Aihara, Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli, Neural Computation 19 (12) (2007) 3335–3355.

[14] T. Rohe, U. Noppeney, Cortical hierarchies perform Bayesian causal inference in multisensory perception, PLoS Biology 13 (2) (2015) e1002073.

[15] C. Kayser, L. Shams, Multisensory causal inference in the brain, PLoS Biology 13 (2) (2015) e1002075.

[16] J. J. Clark, A. L. Yuille, Data fusion for sensory information processing systems, Vol. 105, Springer Science & Business Media, 2013.

[17] J. Trommershauser, K. Kording, M. S. Landy, Sensory cue integration, Oxford University Press, 2011.

[18] B. Rowland, T. Stanford, B. Stein, A Bayesian model unifies multisensory spatial localization with the physiological properties of the superior colliculus, Experimental Brain Research 180 (1) (2007) 153–161.

[19] J. G. Park, S. Jo, Approximate bayesian mlp regularization for regression in the presence of noise, Neural Networks 83 (2016) 75–85.

[20] K. Takiyama, Bayesian geodesic path for human motor control, Neural Networks 93 (2017) 137–142.

[21] T. J. Anastasio, P. E. Patton, K. Belkacem-Boussaid, Using bayes' rule to model multisensory enhancement in the superior colliculus, Neural Computation 12 (5) (2000) 1165–1187.

[22] J. Beck, W. Ma, P. Latham, A. Pouget, Probabilistic population codes and the exponential family of distributions, Progress in Brain Research 165 (2007) 509–519.

[23] W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, Bayesian inference with probabilistic population codes, Nature Neuroscience 9 (11) (2006) 1432–1438.

[24] W. J. Ma, J. M. Beck, A. Pouget, Spiking networks for Bayesian inference and choice, Current Opinion in Neurobiology 18 (2) (2008) 217 – 222.

[25] B. J. Fischer, J. L. Peña, Owl's behavior and neural representation predicted by bayesian inference, Nature neuroscience 14 (8) (2011) 1061.

[26] F. Cazettes, B. J. Fischer, J. L. Pena, Spatial cue reliability drives frequency tuning in the barn owl's midbrain, Elife 3.

[27] F. Cazettes, B. J. Fischer, J. L. Peña, Cue reliability represented in the shape of tuning curves in the owl's sound localization system, Journal of Neuroscience 36 (7) (2016) 2101–2110.

[28] W. J. Ma, M. Rahmati, Towards a neural implementation of causal inference in cue combination, Multisensory Research 26 (2013) 159–176.

[29] W. J. Ma, V. Navalpakkam, J. M. Beck, R. Van Den Berg, A. Pouget, Behavior and neural basis of near-optimal visual search, Nature Neuroscience 14 (6) (2011) 783–790.

[30] Z. Yu, F. Chen, J. Dong, Q. Dai, Sampling-based causal inference in cue combination and its neural implementation, Neurocomputing 175 (2016) 155–165.

[31] C. Cuppini, L. Shams, E. Magosso, M. Ursino, A biologically inspired neurocomputational model for audiovisual integration and causal inference, European Journal of Neuroscience 46 (9) (2017) 2481–2498.

[32] C. Cuppini, E. Magosso, N. Bolognini, G. Vallar, M. Ursino, A neurocomputational analysis of the sound-induced flash illusion, Neuroimage 92 (10) (2014) 248–266.

[33] M. Ursino, A. Crisafulli, G. di Pellegrino, E. Magosso, C. Cuppini, Development of a bayesian estimator for audio-visual integration: A neurocomputational study, Frontiers in computational neuroscience 11 (2017) 89.

[34] M. Ursino, C. Cuppini, E. Magosso, Multisensory bayesian inference depends on synapse maturation during training: theoretical analysis and neural modeling implementation, Neural computation 29 (3) (2017) 735–782.

[35] A. Pouget, J. M. Beck, W. J. Ma, P. E. Latham, Probabilistic brains: knowns and unknowns, Nature Neuroscience 16 (9) (2013) 1170–1178, 13.

[36] M. Carandini, D. J. Heeger, Normalization as a canonical neural computation, Nature Reviews Neuroscience 13 (1) (2012) 51–62.

[37] D. J. Heeger, Normalization of cell responses in cat striate cortex, Visual Neuroscience 9 (2) (1992) 181–197.

[38] C. R. Fetsch, A. Pouget, G. C. DeAngelis, D. E. Angelaki, Neural correlates of reliability-based cue weighting during multisensory integration, Nature Neuroscience 15 (1) (2012) 146–154.

[39] J. M. Fellous, P. H. Tiesinga, P. J. Thomas, T. J. Sejnowski, Discovering spike patterns in neuronal responses, J. Neurosci. 24 (12) (2004) 2989–3001.

[40] L. Shi, T. L. Griffiths, Neural implementation of hierarchical Bayesian inference by importance sampling, in: Advances in neural information processing systems, 2009, pp. 1669–1677.

[41] T. S. Lee, D. Mumford, Hierarchical Bayesian inference in the visual cortex, Journal of the Optical Society of America A 20 (7) (2003) 1434–1448, 14.

[42] R. S. Zemel, P. Dayan, A. Pouget, Probabilistic interpretation of population codes, Neural computation 10 (2) (1998) 403–430.

[43] R. L. De Valois, E. W. Yund, N. Hepler, The orientation and direction selectivity of cells in macaque visual cortex, Vision Research 22 (5) (1982) 531–544.

[44] D. M. Coppola, L. E. White, D. Fitzpatrick, D. Purves, Unequal representation of cardinal and oblique contours in ferret visual cortex, Proceedings of the National Academy of Sciences 95 (5) (1998) 2621–2623.

[45] C. S. Furmanski, S. A. Engel, An oblique effect in human primary visual cortex, Nature Neuroscience 3 (6) (2000) 535.

[46] S. W. Kuffler, Discharge patterns and functional organization of mammalian retina, Journal of neurophysiology 16 (1) (1953) 37–68.

[47] H. K. Hartline, H. G. Wagner, F. Ratliff, Inhibition in the eye of limulus, The Journal of general physiology 39 (5) (1956) 651–673.

[48] S. R. Olsen, V. Bhandawat, R. I. Wilson, Divisive normalization in olfactory population codes, Neuron 66 (2) (2010) 287–299.

[49] O. Brian, D. R. Wozny, S. Ladan, Biases in visual, auditory, and audiovisual perception of space, Plos Computational Biology 11 (12) (2015) e1004649.

[50] M. T. Wallace, G. E. Roberson, W. D. Hairston, B. E. Stein, J. W. Vaughan, J. A. Schirillo, Unifying multisensory signals across time and space., Experimental Brain Research 158 (2) (2004) 252–258.

[51] T. Rohe, U. Noppeney, Sensory reliability shapes perceptual inference via two mechanisms., Journal of Vision 15 (5) (2015) 22.

[52] D. Kappel, S. Habenschuss, R. Legenstein, W. Maass, Network plasticity as Bayesian inference, PLoS Computational Biology 11 (11) (2015) e1004485.

[53] L. Buesing, J. Bill, B. Nessler, W. Maass, Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons, PLoS computational biology 7 (11) (2011) e1002211.

[54] D. Pecevski, W. Maass, Learning probabilistic inference through spike-timing-dependent plasticity, eneuro 3 (2) (2016) ENEURO–0048.

[55] D. Kappel, R. Legenstein, S. Habenschuss, M. Hsieh, W. Maass, A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning, eNeuro 5 (2) (2018) ENEURO–0301.