博士論文（要約）

論文題目　Unconscious Inference in Neural Networks:
Electrophysiology and Learning Theory
（神経回路網における無意識的推論 : 電気生理と学習理論）

氏　　名　磯村　拓哉

Contents

Chapter 1
Introduction

Glossary of expressions

Expression	Description
Generative model	A model that generates the external world dynamics
Internal model	A model in the brain that mimics a generative model of the external world
LTP	Long-term potentiation
LTD	Long-term depression
STDP	Spike-timing dependent plasticity
DCM	Dynamic causal modeling
DEM	Dynamic expectation maximization

1.1 Background

How do people perceive the dynamics of the external world? One hypothesis, the so-called internal model hypothesis [Dayan et al, 1995; Friston, 2006, 2008, 2010; George, Hawkins, 2009; Bastos et al, 2012], states that people reconstruct a model of the external world in their brains through sensory inputs. This internal model helps people infer hidden causes and predict future inputs automatically; in other words, this process happens unconsciously. For example, a songbird can predict a subsequent note in another bird's song by constructing an internal model that mimics the generative model of the bird song [Friston, Kiebel, 2009]. The 19th century physicist/physiologist von Helmholtz hypothesized that, in order to achieve perception, humans are constantly and unconsciously inferring the generative model of the dynamics of the external world. This phenomenon was termed 'unconscious inference' [Helmholtz, Southall, 2005]. A part of unconscious inference has been mathematically modeled under the internal model hypothesis with an existing machine learning model, the Helmholtz machine [Dayan et al, 1995]. However, it is not in a form that can be implemented using actual neural networks. Thus, how actual neural networks implement unconscious inference is not understood. Therefore, the criticism that Helmholtz's theory lacks an explanation of the neural mechanisms (a mechanism theory) is understandable.

Nevertheless, Helmholtz's theory appears intuitively correct as a functional theory of the brain. Therefore, I would like to start the present study under the assumption that brain functions, particularly the higher cognitive functions of the cerebral cortex, can be explained through a machine learning algorithm constructing the internal model [Dayan et al, 1995; Friston, 2006, 2008, 2010; George, Hawkins, 2009; Bastos et al, 2012]. The major goal of the present study is to discover a new machine learning algorithm that is physiologically valid and possesses an expression ability equal to or greater than the existing Helmholtz machine on the basis of experimental observations using actual neural networks. Furthermore, the study must supplement the mechanistic aspects of Helmholtz's theory through that discovery. First, I will introduce a history of the unconscious inference theory, and then describe the purposes and approaches of this thesis.

Learning and memory

Learning and memory are crucial for survival in animals. Memory is defined as the mechanism through which past experiences alter present behavior [Gazzaniga, 2004], and this link between past experiences and present behavior is assumed to be reflected in the physical and biochemical changes in the brain, i.e., as engrams or memory traces. Learning is referred to as the process of obtaining memory.

Recent studies reported that the stimulation of specific neuronal groups related to specific memories could recall and/or rewrite rodent memories in the dentate gyrus [Liu et al., 2012; Ramirez et al., 2013]. Extensive research has focused on learning and memory in not only mammalian systems, but also simpler organisms such as the sea slug *Aplysia* [Abbott, Kandel, 2012].

Synaptic plasticity

Synaptic plasticity is referred to as changes in synaptic strengths. Activity-dependent synaptic plasticity governs the dynamics of synaptic connections and is believed to be a mechanism mediating learning and memory [Bear et al, 2007]. Donald O. Hebb hypothesized that through learning, memories are stored in the brain in the form of networks of neurons, called cell assemblies [Hebb, 1949]; furthermore, he believed that these networks come to represent specific objects and concepts. Hebb further proposed a cellular mechanism of memory formation, known today as Hebbian learning or Hebbian plasticity. This mechanism is best captured in the expression "cells that fire together, wire together."

Spike-timing dependent plasticity (STDP) is an experimentally observed form of Hebbian plasticity [Markram et al, 1997; Bi, Poo, 1998], which is reviewed in [Markram et al, 2011; Feldman, 2012]. In STDP, when the post-synaptic neuron fires immediately after the pre-synaptic neuron firing, the long-term potentiation (LTP) [Matsuzaki et al, 2004; Harvey, Svoboda, 2004] occurs at the connection from the pre- to post-synaptic neuron. In contrast, when the post-synaptic neuron fires immediately before the pre-synaptic neuron does, the long-term depression (LTD) [Zhou et al, 2004] occurs at the connection. Both LTP and LTD involve N-methyl-D-aspartic acid (NMDA)-receptor activity, in which the depolarization of the post-synaptic neuron relieves the NMDA receptors' magnesium block and enables Ca^{2+} entry [Nowak et al,

1984]. Intracellular Ca^{2+} concentration is a roll of switching LTP or LTD, which is regulated by the timing and order of glutamate release from the pre-synaptic terminals and the post-synaptic neuron's depolarization. As the occurrence of LTP or LTD is determined by the order of pre- and post-synaptic neurons' activity, STDP plays a role of increasing the causal relationship between the pre- and post-synaptic neurons. Since the first model of STDP was proposed by [Song et al, 2000], STDP has been modeled using several different equations [Clopath et al, 2010; Gilson, Fukai, 2011].

Unsupervised learning

Sensory perception constitutes complex responses of the brain to sensory input signals. For example, the visual cortex can distinguish objects from their background [DiCarlo et al, 2012], while the auditory cortex can recognize a certain sound in a noisy place with high sensitivity, a phenomenon known as the cocktail party effect [Bronkhorst, 2000; Brown et al, 2001; Mesgarani, Chang, 2012]. Animals have acquired these perceptual abilities without supervision, which is referred to as unsupervised learning [Dayan, Abbott, 2001; Kistler, Gerstner, 2002; Bishop, 2006]. Unsupervised learning is defined as the learning that happens in the absence of a teacher or supervisor; it is achieved through adaptation to experienced environments, which is necessary for cognitive functions. An understanding of the physiological mechanisms that mediate unsupervised learning, or implicit learning, is fundamental to augmenting our knowledge of information processing in the brain. Many researchers have focused their attention on the study of unsupervised learning. However, the human brain has more than one hundred billion neurons, with countless complex connections between them; thus, the physiological mechanisms of unsupervised remain largely unknown.

Theoretical studies in neuroscience

Theoreticians have proposed various models of learning and memory, including models for associative memory, infomax-based learning in the sensory cortex, motor learning in the cerebellum, and reinforcement learning in the striatum [Dayan, Abbott, 2001; Kistler, Gerstner, 2002]. Principal component analysis (PCA) [Oja, 1982] and independent component analysis (ICA) [Bell, Sejnowski, 1995, 1997] are represented using firing rate neuron models and can separate inputs into their individual components. Spiking neuron models show partial learning ability [Clopath et al, 2010; Gilson, Fukai,

2011; Gilson et al, 2012].

Large-scale simulations of neural networks have been achieved using computational approaches to reproduce the dynamics and functions of the brain [Markram, 2006; Izhikevich et al., 2004; Izhikevich, Edelman, 2008; Eliasmith et al, 2012] and to create artificial intelligence capable of performing cognitive tasks. Impressively, one such simulation using the large-scale neural network performed several types of cognitive tasks, including image recognition, reinforcement learning, and working memory [Eliasmith et al, 2012]. Recent progress in deep learning is also remarkable in considering how the brain exhibits higher cognitive functions [Hinton, Salakhutdinov, 2006; Baccouche et al, 2011; Le et al, 2012; LeCun et al, 2015; Lotter et al, 2016].

The brain as an inference machine

Inference means to guess unknown matters based on known facts or certain observations. In other words, inference is a process to draw conclusions through reasoning and estimation. In the ordinary sense of the word, inference is an act of the conscious mind, where consciousness is often considered as a state of self-awareness. Although the importance of consciousness for human cognition is obvious, however, it is widely known that most cognitive processes occur under the unconscious mind.

Hermann von Helmholtz, a 19th century physicist/physiologist, coined the word 'unconscious inference.' In his textbook, he described that

> *"The psychic activities that lead us to infer that there in front of us at a certain place there is a certain object of a certain character, are generally not conscious activities, but unconscious ones. In their result they are equivalent to a conclusion, to the extent that the observed action on our senses enables us to form an idea as to the possible cause of this action."* [Helmholtz, Southall, 2005]

Note that a word 'conclusion' is used as the meaning of inference. In this manner, Helmholtz noticed that the perception often requires inference by the unconscious mind. According to him, an important difference between conscious inference and unconscious inference is whether a conscious knowledge is involved in the process. For example, when an astronomer computes the positions of the stars in space or their distances based on the perspective images he has had at various times and from different parts of the orbit of the earth, he performs conscious inference, which is

"based on a conscious knowledge of the laws of optics. In contrast, in the ordinary acts of vision, this knowledge of optics is lacking" [Helmholtz, Southall, 2005]. Thus, the letter process is performed by the unconscious mind. In spite of such a difference, there is no doubt in the similarity between the results of conscious inference and unconscious inference. Therefore, similar to conscious inference, unconscious inference is crucial for cognitive processes under the unconscious mind to estimate the overall picture from partial observations.

History of unconscious inference theory

As described above, the word 'unconscious inference' was coined by Hermann von Helmholtz (Fig 1-1). He realized that human sensation is not precise; therefore, detailed information should be inferred by the unconscious mind to obtain a precise sensation [Helmholtz, Southall, 2005]. Helmholtz hypothesized that the brain continuously estimates and predicts the dynamics of the external world. Law of prägnanz in Gestalt psychology also indicates that people perceive objects that are close (or similar) to each other as forming a group, which is a kind of inference by the unconscious mind.

In the 1990s, Peter Dayan developed the first machine learning model of unconscious inference, called the Helmholtz machine [Dayan et al, 1995]. He noticed that if animals have a model of the external world in their brain and they continuously optimized its parameters, they can infer the external world. This is termed as the internal model hypothesis. Thus, "*perception is equated with the optimisation or inversion of this internal model, to explain sensory input*" [Friston, Kiebel, 2009]. Let us consider an example of an internal model in the songbird brain, in which a listener bird listens to a singer bird's song, and the singer bird generates the song using several brain regions and produces an output song using the vocal cords. It is known that a generative model of the songbird is given by a two-layered Lorentz attractor [Laje, Mindlin, 2002]. If the listener bird has an internal model of birdsong generation and optimizes its states and parameters, the bird can infer the dynamics of the song and predict what the next note will be [Friston, Kiebel, 2009].

In the 2000s, Karl J. Friston proposed a mathematical foundation of unconscious inference based on the Helmholtz machine, called the free-energy principle [Friston, 2006, 2008, 2010], which is a strong candidate for a unified theory of higher brain functions. He believes that the principle will provide a unified framework of higher

brain functions including perceptual learning [Friston, 2008], reinforcement learning [Reynolds et al, 2001], motor learning [Kilner et al, 2007; Friston et al, 2011], communication [Friston, Frith, 2015a, 2015b], emotion, mental disorders [Fletcher, Frith, 2009; Friston et al, 2014], and evolution. Here, a surprise of input is considered as the criterion of unpredictability.

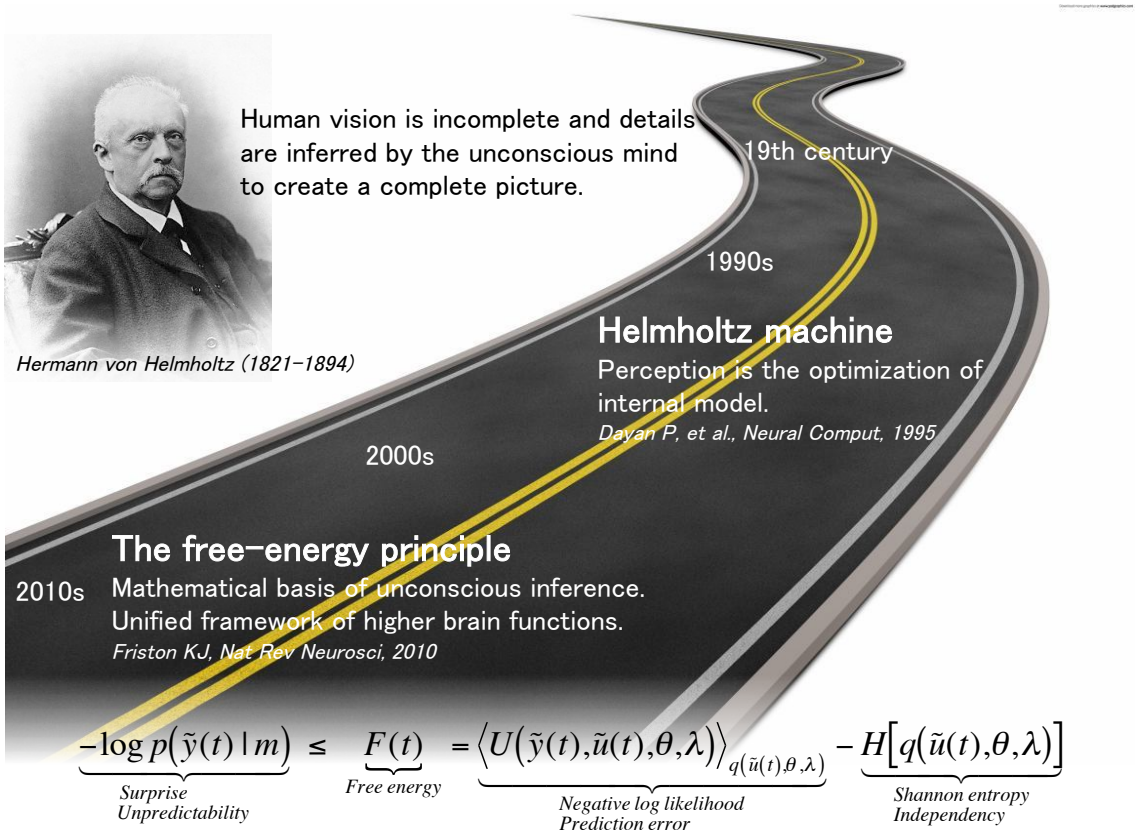

Figure 1-1. History of unconscious inference. The photograph of Helmholtz is reprinted from Wikipedia. Input surprise $-\log p(y(t)|\, m)$ is defined by the rarity of inputs, e.g., when you see a chicken flying in the sky, the surprise of the visual input is very high. The goal of the free-energy principle is to minimize the input surprise.

1.2 The free-energy principle

Information theory

Information is defined as the negative log of probability. If I suppose that $p(s)$ is the probability of a given sensory input, the information in the sensory input is given by $-\log p(s)$ [nat], where nat is a unit of information (1 nat = 1.4427 bits). Here, $-\log p(s)$ is termed as the 'surprise' of the sensory input. For example, a visual input such as that of a chicken flying across the sky has a high surprise value since we have never seen such a scene. The expectation of surprise over $p(s)$ gives the Shannon entropy $H[p(s)] = \langle -\log p(s) \rangle_{p(s)}$ [nat] [Bishop, 2006]. Note that $\langle \bullet \rangle_{p(s)}$ refers to as the expectation over $p(s)$, $\langle \bullet \rangle_{p(s)} = \int \bullet \, p(s) ds$. In the 20th century, Schrödinger assumed that living things minimize the entropy in their body in order to survive [Schrödinger, 1992]. In other words, living things minimize the amount of entropy received from the external world, which is consistent with the minimization of $H[p(s)]$. Indeed, from the viewpoint of self-organization, the entropy reduction for maintenance of life and that for perception and recognition can be considered in a unified manner [Friston, 2013].

The change from a system where s could take two states with the same probability to a system where s could take only one state deterministically decreases 1 bit of entropy. Thus, the brain memorizes the 1-bit information. In other words, the brain state corresponds to 1 bit of the external world state. For the continuous system, I assume a constraint to avoid divergence; I will refer to this constraint as energy. Energy should have a unit of information; the information loss increases if a state goes away from the energy landscape.

Mathematically, the mutual information between the brain and the external world states is defined by $I[\varphi, \vartheta] = H[p_\varphi(\varphi)] + H[p_\vartheta(\vartheta)] - H[p(\varphi, \vartheta)]$, where φ is the brain state and ϑ is the external world state [Bishop, 2006]. Note that $p(\varphi, \vartheta)$ is the joint probability of φ and ϑ, and $p_\varphi(\varphi)$ and $p_\vartheta(\vartheta)$ are their marginal distributions. If the brain state is completely independent of the external world state, $I[\varphi, \vartheta] = 0$ holds. In contrast, if the brain state represents only the external state, $H[p(\varphi, \vartheta)] = H[p_\vartheta(\vartheta)] \geq H[p_\varphi(\varphi)]$ and consequently $I[\varphi, \vartheta] = H[p_\varphi(\varphi)]$ holds. In this manner, the information about the external world stored in the brain is described using $I[\varphi, \vartheta]$. However, the following

requirement becomes obvious:

Requirement 1 (unsupervised):

Information that the brain can access consists only of the sensory input.

Thus, animals have to increase $I[\varphi, \vartheta]$ without the knowledge of ϑ since animals often have difficulty to observe ϑ directly; thus, I will refer to ϑ as hidden states. Accordingly, animals might use $D_{KL}[p_\varphi(\vartheta)|| p_\vartheta(\vartheta)]$ and $-\log p(s)$, where s is the sensory input, instead of using $I[\varphi, \vartheta]$ directly to recognize the external world. Indeed, if I assume κ is the inverse of the signal-noise ratio, I originally find that the relationship of $I[\varphi, \vartheta] \approx \kappa^{-2}(\kappa N/2 - D_{KL}[p_\varphi(\vartheta)|| p_\vartheta(\vartheta)])$ [nat] holds. See Supplementary Information S1.1 for derivation details. In this equation, $p_\varphi(\vartheta)$ is referred to as the recognition density (or the posterior) which expresses the internal model, while $p_\vartheta(\vartheta)$ is the prior which expresses the prior knowledge regarding the hidden states of the world. Both $p_\varphi(\vartheta)$ and $p_\vartheta(\vartheta)$ are stored in the brain, and ϑ are updated through sensory inputs s. Thus, the equation describes that mutual information between the brain and world states can be approximated through sensory inputs. Moreover, a pair containing an agent and an environment can be considered as a kind of thermal bath from a viewpoint of physics. Thus, it is the physical nature of the agent due to which it minimizes the Helmholtz free energy.

The free-energy principle

In the beginning of the 21st century, Friston developed a mathematical foundation of unconscious inference, called the free-energy principle [Friston, 2006, 2008, 2010]. He proposed that the principle will be a unified theory of higher brain functions including perception [Friston, 2008], motor control [Kilner et al, 2007; Friston et al, 2011], reward related learning (conditioning) [Reynolds et al, 2001], social interaction [Friston, Frith, 2015a, 2015b], and even evolution. Each of these functions is fully described by the unified rule, namely the input surprise minimization. First, the goal is the minimization of surprise in the sensory input s, given model m, $-\log p(s|\, m)$, where model m refers to as the prior knowledge of the external world dynamical model [Friston, 2006]. Thus, the principle hypothesizes that animals minimize the input surprise to optimize their perception and behavior. Note that $\langle -\log p(s|\, m)\rangle_{p(s)}$ is always larger than or equal to $\langle -\log p(s)\rangle_{p(s)}$ because of the non-negativity of the

Kullback-Leibler divergence $D_{KL}[p(s)\| p(s|\ m)] \geq 0$ [Bishop, 2006], where $p(s)$ is the true probability density of the sensory input (Fig 1-2).

Since s is generated by the external world generative model, it is better to consider $p(s, \vartheta|\ m)$ for better inference, where ϑ is the external world invisible (hidden) state (including hidden variables, parameters, and hyper-parameters). To deal with the hidden state ϑ, one strategy could be to develop the internal model in the brain [Dayan et al, 1995; Friston, 2006, 2008, 2010; George, Hawkins, 2009; Bastos et al, 2012], where the internal model is defined as an estimation model of the generation of input signals established through unsupervised learning. Unfortunately, $-\log p(s|\ m) = -\log \int p(s, \vartheta|\ m)d\vartheta$, which is the cost function of restricted Boltzmann machine (RBM) [Smolensky, 1986; Hinton, 2002], is intractable for animals, since they have to deal with the integral of $p(s, \vartheta|\ m)$ placed in the logarithm function.

The principle hypothesizes that animals calculate an upper bound of $-\log p(s|\ m)$ instead, that is tractable for animals. In this manner, the free energy F is defined by

$$F(s, q(\vartheta); t) = -\log p(s|\ m) + D_{KL}[q(\vartheta) \| p(\vartheta|s, m)], \tag{1.1}$$

where $q(\vartheta)$ is the recognition density—the probability density of the internal model in the brain [Friston, 2008, 2010]. Formally speaking, $q(\vartheta)$ is obtained by substituting $\varphi = \vartheta$ into the probability density of the brain state given input s, $q(\vartheta) = p_\varphi(\varphi|\ s)|_{\varphi=\vartheta} = p_\varphi(\vartheta|\ s)$. Due to the non-negativity of the Kullback-Leibler divergence $D_{KL}[q(\vartheta) \| p(\vartheta|s, m)] \geq 0$, F provides an upper bound of $-\log p(s|\ m)$ (Fig 1-2). Thus, $D_{KL}[q(\vartheta) \| p(\vartheta|s, m)]$ indicates the difference between actual probabilities of hidden states $p(\vartheta|s, m)$ and their expected probabilities $q(\vartheta)$.

Based on the definition of the Kullback-Leibler divergence ($D_{KL}[q(\vartheta) \| p(\vartheta|s, m)] = \langle \log q(\vartheta) - \log p(\vartheta|\ s, m)\rangle_{q(\vartheta)}$), the free energy is transformed to $F(s, q(\vartheta); t) = \langle -\log p(s, \vartheta|\ m) + \log q(\vartheta)\rangle_{q(\vartheta)}$. Here, I define the first term as Gibbs energy (or the internal energy) $G(s, \vartheta; t) = -\log p(s, \vartheta|\ m)$, which refers to as the amplitude of the prediction error at a given moment [Friston, 2008, 2010]. The second term is the Shannon entropy, $H[q(\vartheta)] = \langle -\log q(\vartheta)\rangle_{q(\vartheta)}$. Therefore, I obtain

$$F(s, q(\vartheta); t) = \langle G(s, \vartheta; t)\rangle_{q(\vartheta)} - H[q(\vartheta)]. \tag{1.2}$$

The first term of Eq. (1.2), the expectation of G over $q(\vartheta)$, represents the error of estimation. Optimization of $\langle G(s, \vartheta; t)\rangle_{q(\vartheta)}$ is the same as the maximum a posteriori

(MAP) estimation (or the maximum likelihood estimation in the case that a prior has a uniform distribution) [Bishop, 2006]. Thus, the first term represents the homeostasis of a biological system allowing it to adapt to its environment, while the second term represents the diversity (or exploration) of the inner states. To minimize the free energy with a condition of low background noise, it is necessary to maximize $H[q(\vartheta)]$, and thus maximize the independence of the inner state. As indicated by Jaynes, the maximization of entropy $H[q(\vartheta)]$ is crucial to biological systems [Jaynes, 1957a, 1957b]. Specifically, the maximization of $H[q(\vartheta)]$ is essential for blind source separation (the inference of hidden causes) because the optimal parameters that minimize the prediction error are not always determined identically, and the MAP estimation is not always found with parameters that separate the sensory inputs into independent hidden sources. Therefore, free-energy minimization is the rule to simultaneously minimize the prediction error and maximize the independence of the inner states.

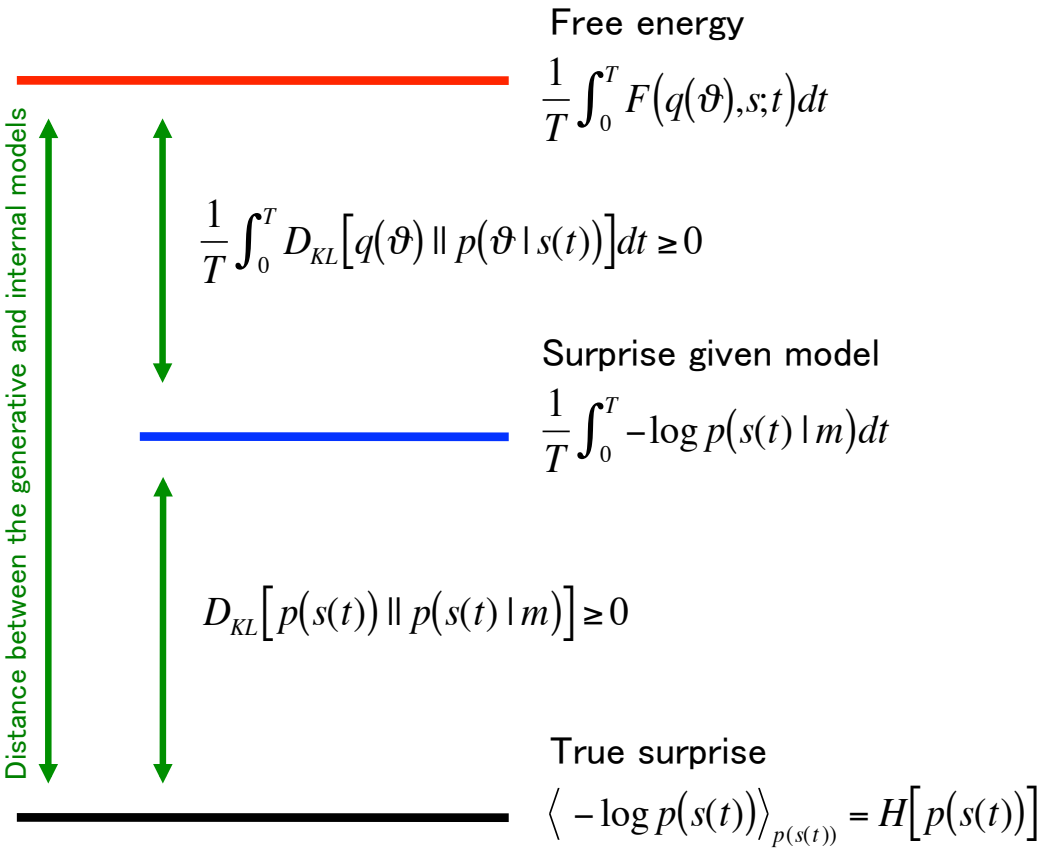

Figure 1-2. A schematic image of energy level. Because of the non-negativity of the Kullback-Leibler divergence, $\langle -\log p(s|\, m)\rangle_{p(s)}$ is always larger than or equal to $\langle -\log p(s)\rangle_{p(s)}$ and F provides an upper bound of $-\log p(s|\, m)$.

Hypothesized generative and internal models

Under the free-energy principle, perception and learning indicate the optimization of the internal model that mimics the generative model of the external world dynamics. I suppose the external world generative model to be described as

$$\begin{aligned}
&\dot{x} = f(x, v; \theta) + w, \\
&s = g(x, v; \theta) + z, \\
&v_i \sim \text{i.i.d.} \sim p_v(v_i), \\
&w \sim N[w; 0, \Sigma^w(\lambda)] \text{ where } \langle w(t)w(t')^T \rangle = \Sigma^w \delta(t-t'), \\
&z \sim N[z; 0, \Sigma^z(\lambda)] \text{ where } \langle z(t)z(t')^T \rangle = \Sigma^z \delta(t-t'),
\end{aligned} \tag{1.3}$$

where s is the sensory input, x is the hidden state, v is the source, w and z are the noises, θ is a set of parameters, λ is a set of hyperparameters, and f and g are nonlinear functions. Sources v and noises w and z are supposed to be random variables that follow $p_v(v) = \prod_i p_v(v_i)$, $p_w(w) = N[w; 0, \Sigma^w(\lambda)]$, and $p_v(v) = N[v; 0, \Sigma^v(\lambda)]$, respectively. Such an approach considering the generative and internal models is referred to as dynamic causal modeling (DCM) [Friston, 2008]. I define errors by $\varepsilon^x = Dx - f$ and $\varepsilon^v = s - g$. A set of x and v will be referred to as $u = (x, v)$. Under the assumption of such a generative model, the first and second terms of free energy $\langle G(s, \vartheta) \rangle_{q(\vartheta)}$ and $H[q(\vartheta)]$ can be explicitly calculated (see S1.2 for details), where ϑ is a set of hidden states $\vartheta = \{u, \theta, \lambda\}$. Moreover, I hypothesize that recognition densities are represented as $q(\vartheta) = q(u)\, q(\theta)\, q(\lambda)$ (the mean field approximation) and they respectively follow $q(u) = N[u; \mu, C^u]$, $q(\theta) = N[\theta; \boldsymbol{\theta}, C^\theta]$, and $q(\lambda) = N[\lambda; \boldsymbol{\lambda}, C^\lambda]$ (Laplace approximation). Note that μ, $\boldsymbol{\theta}$, and $\boldsymbol{\lambda}$ are the expectations of u, θ, and λ, respectively, and C^u, C^θ, and C^λ are their covariance matrices. This is a mathematical representation of the internal model.

The free-energy principle predicts that μ, $\boldsymbol{\theta}$, and $\boldsymbol{\lambda}$ are established in a manner to minimize the free energy, such that $\mu = \arg\min F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})$, $\boldsymbol{\theta} = \arg\min F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})$, and $\boldsymbol{\lambda} = \arg\min F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})$. According to the gradient descent scheme, I obtain update rules for μ, $\boldsymbol{\theta}$, and $\boldsymbol{\lambda}$ as the following:

$$\begin{aligned}
&\dot{\mu} \propto D\mu - V^u(u)_u|_{\mu=u} \approx D\mu - F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})_\mu, \\
&\dot{\boldsymbol{\theta}} \propto -V^\theta(\theta)_\theta|_{\boldsymbol{\theta}=\theta} \approx -F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})_{\boldsymbol{\theta}},
\end{aligned}$$

$$\dot{\boldsymbol{\lambda}} \propto -V^{\lambda}(\lambda)_{\lambda}|_{\lambda=\lambda} \approx -F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})_{\boldsymbol{\lambda}}, \tag{1.4}$$

where $V^{u}(u) = \langle U(s, \vartheta)\rangle_{q(\theta,\lambda)}$, $V^{\theta}(\theta) = \langle U(s, \vartheta)\rangle_{q(u,\lambda)}$, and $V^{\lambda}(\lambda) = \langle U(s, \vartheta)\rangle_{q(u,\theta)}$ are the variational energies and are approximately equal to $F(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})$ except the constant term. $D\mu$ indicates the original trajectory without perturbation, while μ is the change in μ's trajectory after the perturbation by $V^{u}(u)_{u}|_{\mu=u}$. This procedure for updating states, parameters, and hyper-parameters is referred to as dynamic expectation maximization (DEM) [Friston, 2008].

The free-energy principle, the internal model hypothesis, and neurophysiology

The relationship between cortical microcircuits and predictive coding model has been investigated [Bastos et al, 2012]. The predictive coding model is consistent with previous biological knowledge and proposes the function of these microcircuits. Moreover, it is known that spontaneous prior activity of a visual area learns the properties of natural pictures [Berkes et al, 2011]. A recent study showed that cortical neurons in rodents code hidden states in accordance with the Bayesian brain hypothesis [Funamizu et al, 2016]. These results suggest that the free-energy principle is plausible as the theory of higher brain functions.

1.3 Problems of the free-energy principle

The free-energy principle is deterministically a good theory from a psychological point of view (or information, theoretical, and engineering points of view). From the free-energy principle, it is hypothesized that (Hypothesis 1) all the learning rules in the brain can be defined as a derivative of a common cost function, which is referred to as the free energy F [Friston, 2008, 2010]. Accordingly, the free-energy principle can unify learning mechanisms of various cognitive functions including pattern memorization, probability learning (Pavlovian learning), and dynamics (sequence) learning. However, in order for the free-energy principle to be a physiologically plausible theory in the brain, the principle needs to satisfy certain physiological requirements and verify the applicability to complicated and realistic situations. I will point out three major problems in the free-energy principle as the followings:

[1] Problems of Fristonian neurophysiology: a lack of physiological evidence

Although the free-energy principle is a simple and plausible rule from an information theory perspective, there are several problems with it from a biological point of view. First, electrophysiological data that elucidate the neural and synaptic bases of this theory are lacking.

Fiorillo criticized the free-energy principle for the paucity of electrophysiological evidence at the levels of microcircuits, neurons, and synapses [Fiorillo, 2010]. The free-energy principle can explain the functional aspects of higher brain functions. Although some laminar-specific structures in the cortex are consistent with predictive coding based on the free-energy principle [Bastos et al, 2012], there is little physiological evidence given the difficulty of recording neuron-neuron interactions *in vivo*. Therefore, it is necessary to directly observe the learning (or self-organizing) processes of unconscious inference using actual neural networks.

Ambiguity in the structures of the internal model (the recognition model) is an additional problem. It is unclear how neural networks encode sensory information when they obtain new recognition models. The older unsupervised learning models (PCA and ICA) employ the inverse recognition model that learns the inverse of a matrix in the

generative model [Oja, 1982; Bell, Sejnowski, 1995, 1997]. On the other hand, newer models (the sparse coding model, RBM, and the predictive coding model) use the feed-forward recognition model that learns the mixing matrix itself [Olshausen, 1996; Olshausen, Field, 1997; Hinton, Salakhutdinov, 2006; Friston, 2008, 2010]. It is possible that neural networks use both the models for learning, although Friston and colleagues only discussed the feed-forward model [Friston, 2008, 2010; Bastos et al, 2012]. For instance, the activity of actual neural networks does not always converge into an equilibrium state with the input stimulation, since actual neural networks tend to generate spikes synchronously, while the feed-forward models require equilibrium of the inner states. Therefore, the feed-forward model may not be appropriate in the synchronous-input case. It will be necessary to find biologically plausible structures or models of networks that can explain the free-energy principle in actual neural networks.

Moreover, the learning rule hypothesized by the free-energy principle is not plausible either. The learning rule should be explained using the biologically plausible Hebbian plasticity [Hebb, 1949], such as STDP [Markram et al, 1997; Bi, Poo, 1998]. Although the free-energy principle appears to be consistent with various behaviors (at the systems or macro levels) and cortical microcircuits structures (at the circuit or mesoscopic levels), it is debatable whether the free-energy principle can explain the dynamics of neurons and synapses (at the cellular or micro levels). The free-energy principle hypothesizes that the parameters in the internal model are represented by synaptic strengths, which are established through a Hebbian-like update rule [Friston, 2008]. Recent studies report an essential role of GABAergic transmission in modulating Hebbian plasticity to an anti-Hebbian (or STDP to anti-STDP) manner [Hayama et al, 2013; Paille et al, 2013]. Such modulation may be important while considering a biologically plausible learning rule for unconscious inference.

[2] Local learning rule: a physiological constraint on update rules

As described above, the internal model hypothesis explains a functional aspect alone—conventional models have problems that can be implemented by actual neural networks. The neuronal mechanisms by which the neural network implements the internal model are largely unclear. Importantly, neurons communicate through firing (spiking) activity. Therefore, the information that neurons can access only comprises the activities of other neurons connected to them via synapses (local information):

Requirement 2 (locality):

The only information that neurons can access is local information. Here, local information is defined as the firing information of the connected neurons, which the neurons can access through synaptic connections. Therefore, any learning rule employed by neurons must be a local learning rule using only local information.

Unfortunately, the DCM (mainstream kind of modeling based on the free energy principle [Friston, 2008]; see also the previous section) uses a non-local learning rule. However, given this requirement, I newly hypothesize that (Hypothesis 2) any learning rule derived from the free energy must be a three-factor learning rule [Frémaux, Gerstner, 2016], which is a biologically popular local learning rule in the literature of reinforcement learning [Reynolds et al, 2001], and be physiologically implemented by neuromodulated Hebbian plasticity [Pawlak et al, 2010].

[3] Reliability in establishing internal models under complicated environments

Considering that generative models in the real world have nonlinear and hierarchical structures, the theory of unconscious inference has to address them. However, it remains unclear whether and how the internal model can stably identify the nonlinear and hierarchical dynamical systems. Although dynamic causal modeling (DCM) under the free-energy principle [Friston, 2008] has addresses these issues, theoretical studies that guarantee the stability and reliability of solutions are lacking. Indeed, when the cost function has many local minima, an efficient global search is required to avoid local minima and reach a global minimum, while the current studies use a simple gradient descent approach that is weak to local minima [Friston, 2008].

Moreover, in the real world, there are often more than two agents existing simultaneously. However, the free-energy principle has not been successful in addressing the multi-agent problems yet, while it typically consider interactions between an agent and the external world [Kilner et al, 2007; Friston et al, 2011], or between two agents [Friston, Frith, 2015a, 2015b].

Taken all together, I have pointed out three major problems in the free-energy principle; [1] physiological evidence that shows the existence of learning or self-organizing processes under the free-energy principle is lacking; [2] the update rule

must be a biologically plausible local learning rule, while the current rule is non-local; and [3] the unconscious inference theory must be applicable to complicated environments including nonlinear and hierarchical dynamics and interactions between multiple agents.

Accordingly, it is necessary to investigate how actual neural networks infer the dynamical system or the generative model behind the sensory input, and to develop a biologically plausible mathematical algorithm (learning rule) through which the actual neural network might implement the internal model in a manner consistent with the physiological experimental observations.

1.4 Purpose

As described above, theoreticians hypothesize that the brain develops an internal model that mimics the external world generative model to perform unconscious inference according to the input surprise (free energy) minimization. Thus, the hypothesis will provide a unified theory of higher brain functions including perceptual, reinforcement, and motor learning. However, the hypothesis only explains the functional aspects, and the mechanism through which the neural network implements the internal model in a physiologically plausible manner is not understood. Therefore, the neuronal mechanism mediating the implementation of unconscious inference through the modification of synaptic strengths is largely unclear. Therefore, the purposes of this thesis are as follows:

Purpose 1:

To test the hypothesis that "even simple neural networks can perform unconscious inference by developing the internal model through neuromodulated Hebbian plasticity".

Purpose 2:

To develop a local learning algorithm through which neural networks can develop the internal model for the stochastic dynamical system.

Purpose 3:

To demonstrate the usefulness of the multiple internal model for efficient global search and inference of complicated environments.

To examine these hypotheses, I will conduct the following *in vitro* experiments in addition to developing mathematical models.

1.5 Approach

This thesis is divided into 5 chapters. In Chapter 1, I summarized the history of studies of unconscious inference and identify the problems and gaps in the current knowledge. Next, I mentioned the purposes of this thesis.

In Chapter 2, I will examine whether cultural neural networks satisfy the requirements stated earlier. This is a so-called constructive approach, which is bottom-up approach to develop a system in order to understand the mechanism and requirements of the system. Using the constructive approach, it is possible to reconstruct a part of the functions, which I term as 'cognitive-like functions,' within cultured neural networks. Many studies have identified the properties of learning and memory in cultured neural networks. For example, cultured neural networks demonstrate pathway-specific synaptic plasticity induced by local tetanic stimulation [Jimbo et al, 1999], supervised learning and adaptation in response to input signals [Shahaf, Marom, 2001; Eytan et al, 2003], pattern recognition and associative memory [Ruaro et al, 2005], and behaving as logistic gate devices and diodes [Feinerman, Moses, 2006; Feinerman et al, 2008]. Dissociated networks of cultured neurons do not maintain their natural biological structure; however, these previous studies suggest that cultured neurons are capable of learning and memory processes.

In this chapter, I will explore biologically plausible models and rules that establish the internal model using dissociated cultures of the rat cerebral cortex. I will investigate the changes in the evoked responses of neurons to electrical stimulation, particularly focusing on tasks of blind source separation and MAP estimation, both of which are requirements of system identification and generative model inference.

First, I will reproduce the system identification ability in cultured neural networks. Next, I will determine the learning model that is employed, and calculate the free energy values for the system, using the estimated connection strengths from the evoked responses. Finally, I will determine the learning rules and pharmacologically examine how GABAergic input influences the learning processes to investigate the physiological mechanism.

In Chapter 3, I will discuss the learning theory for neural networks. Computational modeling is also an example of a constructive approach. In order to fully understand the nature of the inference, a theory that can explain and predict neural dynamics and behavior is required.

In Section 3.2, I will summarize a local learning rule for ICA that I developed. In Section 3.3, I will develop a novel local learning rule for PCA and ICA. In Section 3.4, I will apply the rule to temporally decompose, memorize, and predict the input sequence. In Sections 3.5 and 3.6, I will apply the learning rule to multi-context processing and nonlinear blind source separation. Finally, I will discuss the relationship between the proposed models and the free-energy principle.

In Chapter 4, I will study the approaches to tackle nonlinear system identification. First, I will calculate the cost for global search using random and local searches. Next, I will calculate the cost while using the crossover search algorithm. Finally, I will propose a multiple internal model to infer the mind of another individual.

In Chapter 5, I will describe the highlights of this thesis.

Chapter 2

Neuronal system identification

Because the contents of Chapter 2 will be published from a scientific journal soon, they will not be published on the internet for five years.

Chapter 3
Local learning rule for unconscious inference

Because the contents of Chapter 3 will be published from a scientific journal soon, they will not be published on the internet for five years.

Chapter 4
Multiple internal model

Because the contents of Chapter 4 will be published from a scientific journal soon, they will not be published on the internet for five years.

Chapter 5
Discussion and conclusion

5.1 Discussion

The free-energy principle is a candidate unified theory of unconscious inference [Friston, 2010]. However, as I pointed out in Section 1.3, three major problems exist, which I repeat in Table 5-1 left. To solve them, I conducted electrophysiological experiments and theoretical studies in Chapters 2–4 (Table 5-1 middle column).

In Chapter 2, I observed that cultured neural networks could perform blind source separation, predictive coding, and stochastic dynamical system identification (Table 5-1 top). I explicitly showed that these learning processes were followed by the free energy reductions as predicted by the free-energy principle. These results provide the first formal evidence of neuronal self-organization under the free-energy principle. Moreover, these learning processes were possibly mediated by GABA-modulated Hebbian plasticity.

In Chapter 3, on the basis of observations in Chapter 2, I developed a local three-factor learning rule, error-gated Hebbian rule (EGHR), which is consistent with GABA-modulated Hebbian plasticity [Nishiyama et al, 2010; Hayama et al, 2013; Paille et al, 2013; Müllner et al, 2015] and heuristically derived from the free-energy principle through an approximation to meet the requirement of the locality (see Section 3.7). I demonstrated that the EGHR can perform PCA, ICA and MAP estimation while a generative has recurrent dynamics, nonlinear transformations, and hierarchical structures, and even when several generative models switch from time to time (Table 5-1 middle row). Taken together, the EGHR can be a biologically more plausible alternative of a currently mainstream non-local update rule under the free-energy principle [Friston, 2008].

In Chapter 4, I showed that the multiple internal model is useful for the efficient global search and the inference of complicated environments (Table 5-1 bottom). In Section 4.1, I estimated the expectation of calculation costs of optimization problems when cost functions can be expanded into a Fourier series and proposed an optimal searching algorithm based on multiple internal models, which enables to enhance the global search efficacy of optimal hidden states and parameters under the free-energy principle. In Section 4.2, I proposed an application of the multiple internal model to

infer multiple generative models or another's mind and solve a self-other distinction problem.

Taken all together, I succeeded to solve these problems and enhance the biological plausibility and the reliability of the free-energy principle.

Table 5-1. Summary of results.

	Problem	**Purpose**	**Result**
[1] Ch 2	Physiological evidence that shows the existence of learning or self-organizing processes under the free-energy principle is lacking.	To examine whether simple neural networks perform unconscious inference by developing the internal model through neuromodulated Hebbian plasticity".	Cultured neural networks could perform inference of dynamical systems, which reduced free energy and might be mediated by GABA-modulated Hebbian rule.
[2] Ch 3	The update rule must be a biologically plausible local learning rule, while the current rule is non-local.	To develop a local learning algorithm through which neural networks can develop the internal model for the stochastic dynamical system.	The proposed local three-factor learning rule, or the EHGR, can perform inference over a wide range of generative models.
[3] Ch 4	The theory must be applicable to complex environments including many local minima and interactions between multiple agents.	To demonstrate the usefulness of the multiple internal model for efficient global search and inference of complex environments.	The multiple internal model can accelerate a speed of a global search, separately infer multiple generative model, and solve a self-other distinction problem.

5.2 Conclusion

Unconscious inference and the free-energy principle are well-established theories in psychology and theoretical neuroscience; however, there has been little physiological evidence that neuronal microcircuits perform them. Main discoveries of this thesis are as follows:

Conclusion 1:

Cultured neural networks can develop the internal model and perform unconscious inference of stochastic dynamical system, including blind source separation, predictive coding, and stochastic dynamical system identification, which reduces free energy as predicted by the free-energy principle. It is suggested that the state-dependent Hebbian plasticity mediated by GABAergic input underlies this learning process.

Conclusion 2:

The proposed local three-factor learning rule, or the EGHR, is heuristically derived from the free-energy principle and can perform unconscious inference of a wide range of stochastic dynamical systems in a biologically plausible manner, which is consistent with my observations as well as recent physiological findings.

Conclusion 3:

The multiple internal model enables to enhance the global search efficacy of optimization problems and to infer multiple generative models. The latter is useful for inferring another's mind and solving a self-other distinction problem.

My results indicate that biological neural circuits perform unconscious inference similar to how machine learning proceeds in computers. These results will not only lead to a better understanding of the nature of learning in the brain, but also suggest that intelligence can emerge with properties common to the biological nervous system and computers. In the future, I plan to further explore the nature of intelligence by using both biological systems and machine learning approaches.

Acknowledgements

I would like to show my greatest appreciation to Prof. Yasuhiko Jimbo who offered continuing support and constant encouragement. I would also like to thank Prof. Kiyoshi Kotani whose comments have helped me.

Special thanks also go to Dr. Taro Toyoizumi, RIKEN Brain Science Institute, whose comments and opinions made enormous contribution to my work throughout the production of this study. I am also grateful to Prof. Karl J. Friston, University College London, for helpful discussions.

I would like to thank Dr. Yutaro Ogawa and Mr. Akihiko Akao for helpful comments to this manuscript. Special thanks also go to other colleagues who gave me invaluable comments and warm encouragements.

I would also like to express my gratitude to my family for their moral support and warm encouragements.

Note that Chapter 2 describes works in Jimbo's laboratory; Chapter 3 describes works in Toyoizumi's laboratory; and Section 4.3 is a work with Dr. Friston.

The responsibility for the final formulation, and any errors that it may concern, are entirely mine.

References

Amari, S. I. (1977). Neural theory of association and concept-formation. Biological cybernetics, 26(3), 175-185.

Amari, S. I., Cichocki, A., & Yang, H. H. (1996). A new learning algorithm for blind signal separation. Advances in neural information processing systems, 757-763.

Amari, S. I., Douglas, S. C., Cichocki, A., & Yang, H. H. (1997, April). Multichannel blind deconvolution and equalization using the natural gradient. In Signal Processing Advances in Wireless Communications, First IEEE Signal Processing Workshop on (pp. 101-104). IEEE.

Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., & Baskurt, A. (2011). Sequential deep learning for human action recognition. In Human Behavior Understanding (pp. 29-39). Springer Berlin Heidelberg.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. Cognition, 21(1), 37–46.

Barth, A. L., & Poulet, J. F. (2012). Experimental evidence for sparse firing in the neocortex. Trends in neurosciences, 35(6), 345–355.

Barto, A.G., Sutton, R.S., & Anderson, C.W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. Systems, Man and Cybernetics, IEEE Transactions on. 13(5), 834–846.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. Neuron, 76(4), 695-711.

Bear, M. F., Connors, B. W., & Paradiso, M. A. (Eds.). (2007). Neuroscience (Vol. 2). Lippincott Williams & Wilkins.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), 1129-1159.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. Vision research, 37(23), 3327-3338.

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. Science, 331(6013), 83-87.

Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of neuroscience, 18(24), 10464-10472.

Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern Recognition and Machine Learning. Springer, New York.

Bolander, T. (2014). Seeing is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic. In ECSI (pp. 87-107).

Brito, C. S., & Gerstner, W. (2016). Nonlinear Hebbian learning as a unifying principle in receptive field formation. arXiv preprint arXiv:1601.00701.

Brown, G. D., Yamada, S., & Sejnowski, T. J. (2001). Independent component analysis at the neural cocktail party. Trends in neurosciences, 24(1), 54-63.

Chiappalone, M., Massobrio, P., & Martinoia, S. (2008). Network plasticity in cortical assemblies. European Journal of Neuroscience, 28(1), 221-237.

Cichocki, A., Karhunen, J., Kasprzak, W., & Vigario, R. (1999). Neural networks for blind separation with unknown number of sources. Neurocomputing, 24(1), 55-93.

Cichocki A, Zdunek R, Phan AH, Amari SI (2009) Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. John Wiley & Sons.

Choi, S., Amari, S., Cichocki, A., & Liu, R. W. (1999). Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels. In Int. Workshop on ICA and BSS (pp. 371-376).

Clopath, C., Büsing, L., Vasilaki, E., & Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nature neuroscience, 13(3), 344-352.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. Neural computation, 7(5), 889-904.

Dayan, P., & Abbott, L. F. (2001). Theoretical neuroscience (Vol. 10, pp. S0306-4522). Cambridge, MA: MIT Press.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. Neuron, 73(3), 415-434.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. Nature Reviews Neuroscience, 11(2), 114-126.

Doya, K. (2002). Metalearning and neuromodulation. Neural Networks. 15(4),

495–506.

Doya, K. (2007). Reinforcement learning: Computational theory and biological mechanisms. HFSP Journal. 1(1), 30–40.

Dranias, M. R., Ju, H., Rajaram, E., & VanDongen, A. M. (2013). Short-term memory in networks of dissociated cortical neurons. The Journal of Neuroscience, 33(5), 1940-1953.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. science, 338(6111), 1202-1205.

Eytan, D., Brenner, N., & Marom, S. (2003). Selective adaptation in networks of cortical neurons. The Journal of neuroscience, 23(28), 9349-9356.

Feinerman, O., & Moses, E. (2006). Transport of information along unidimensional layered networks of dissociated hippocampal neurons and implications for rate coding. The Journal of neuroscience, 26(17), 4526-4534.

Fei-Fei, L., Fergus, R. & Perona, P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. IEEE CVPR 2004, Workshop on Generative-Model Based Vision 2004. 178 (2003).

Feinerman, O., Rotem, A., & Moses, E. (2008). Reliable neuronal logic devices from patterned hippocampal cultures. Nature physics, 4(12), 967-973.

Feldman, D. E. (2012). The spike-timing dependence of plasticity. Neuron, 75(4), 556-571.

Fiete, I. R., Senn, W., Wang, C. Z., & Hahnloser, R. H. (2010). Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron, 65(4), 563-576.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience, 10(1), 48-58.

Florian, R. V. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. Neural Computation, 19(6), 1468-1502.

Fogelson, N., Litvak, V., Peled, A., Fernandez-del-Olmo, M., & Friston, K. (2014). The functional anatomy of schizophrenia: a dynamic causal modeling study of predictive coding. Schizophrenia Research. 158(1), 204–212.

Földiak, P. (1990). Forming sparse representations by local anti-Hebbian learning.

Biological cybernetics, 64(2), 165-170.

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature, 440(7084), 680-683.

Frémaux, N., Sprekeler, H., & Gerstner, W. (2010). Functional requirements for reward-modulated spike-timing-dependent plasticity. The Journal of Neuroscience, 30(40), 13326-13337.

Frémaux, N., & Gerstner, W. (2016). Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules. Frontiers in neural circuits, 9.

Friston, K. J., Frith, C. D., & Frackowiak, R. S. J. (1993). Principal component analysis learning algorithms: a neurobiological analysis. Proceedings of the Royal Society of London B: Biological Sciences, 254(1339), 47-54.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. Journal of Physiology-Paris, 100(1), 70-87.

Friston, K. (2008). Hierarchical models in the brain. PLoS computational biology, 4(11), e1000211.

Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. Neural Networks, 22(8), 1093-1104.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature Reviews Neuroscience, 11(2), 127-138.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. Biological cybernetics, 104(1-2), 137-160.

Friston, K. (2013). Life as we know it. Journal of the Royal Society Interface, 10(86), 20130475.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. The Lancet Psychiatry, 1(2), 148-158.

Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. cortex, 68, 129-143.

Friston, K., & Frith, C. (2015). A duet for one. Consciousness and cognition, 36, 390-405.

Frith, U. (2001). Mind blindness and the brain in autism. Neuron, 32(6), 969-979.

Froemke, R. C., Merzenich, M. M., & Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. Nature, 450(7168), 425-429.

Funamizu, A., Kuhn, B., & Doya, K. (2016). Neural substrate of dynamic Bayesian inference in the cerebral cortex. Nature Neuroscience, 19(12), 1682-1689.

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nature neuroscience, 17(5), 732.

Gazzaniga, M. S. (2004). The cognitive neurosciences. MIT press.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on, (6), 721-741.

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. PLoS computational biology, 5(10), e1000532.

Gerstner, W., & Kistler, W. M. (2002). Spiking neuron models: Single neurons, populations, plasticity. Cambridge university press.

Gilson, M., & Fukai, T. (2011). Stability versus neuronal specialization for STDP: long-tail weight distributions solve the dilemma. PloS one, 6(10), e25339.

Gilson, M., Fukai, T., & Burkitt, A. N. (2012). Spectral analysis of input spike trains by spike-timing-dependent plasticity. PLoS Comput Biol, 8(7), e1002584.

Harvey, C. D., & Svoboda, K. (2007). Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature, 450(7173), 1195-1200.

Hayama, T., Noguchi, J., Watanabe, S., Takahashi, N., Hayashi-Takagi, A., Ellis-Davies, G. C., ... & Kasai, H. (2013). GABA promotes the competitive selection of dendritic spines by controlling local Ca2+ signaling. Nature neuroscience, 16(10), 1409-1416.

Hebb, D. O. (1949). The organization of behavior: A neuropsychological theory (John Wiley & Sons, New York).

von Helmholtz, H., & Southall, J. P. C. (2005). Treatise on physiological optics (Vol. 3). Courier Corporation.

Henneberger, C., Papouin, T., Oliet, S. H., & Rusakov, D. A. (2010). Long-term potentiation depends on release of D-serine from astrocytes. Nature, 463(7278), 232-236.

Hensch, T.K. & Fagiolini, M. Excitatory-inhibitory balance and critical period plasticity in developing visual cortex. Prog. Brain Res. 147, 115–124 (2005).

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate

brain. Frontiers in human neuroscience, 3.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. Neural Computation, 14(8), 1771-1800.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507.

Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. Nature Reviews Neuroscience, 10(9), 647-658.

Insel, T. R. (2010). Rethinking schizophrenia. Nature, 468(7321), 187–193.

Isomura, T., Kotani, K., & Jimbo, Y. (2015). Cultured Cortical Neurons Can Perform Blind Source Separation According to the Free-Energy Principle. PLoS Comput Biol, 11(12), e1004643.

Isomura, T., Sakai, K., Kotani, K., & Jimbo, Y. (2016). Linking Neuromodulated Spike-Timing Dependent Plasticity with the Free-Energy Principle. Neural Computation, 28(9), 1859-1888.

Isomura, T., & Toyoizumi, T. (2016). A Local Learning Rule for Independent Component Analysis. Scientific Reports, 6.

Izhikevich, E. M., Gally, J. A., & Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. Cerebral cortex, 14(8), 933-944.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. Cerebral cortex, 17(10), 2443-2452.

Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. Proceedings of the national academy of sciences, 105(9), 3593-3598.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 304(5667), 78-80.

Jaynes, E. T. (1957). Information theory and statistical mechanics. Physical Rseview, 106(4), 620.

Jaynes, E. T. (1957). Information theory and statistical mechanics. II. Physical Review, 108(2), 171.

Jimbo, Y., Tateno, T., & Robinson, H. P. C. (1999). Simultaneous induction of pathway-specific potentiation and depression in networks of cortical neurons. Biophysical Journal, 76(2), 670-678.

Jimbo, Y., Kasai, N., Torimitsu, K., Tateno, T., & Robinson, H. P. (2003). A system for

MEA-based multisite stimulation. Biomedical Engineering, IEEE Transactions on, 50(2), 241-248.

Johansen, J. P., Diaz-Mataix, L., Hamanaka, H., Ozawa, T., Ycu, E., Koivumaa, J., ... & LeDoux, J. E. (2014). Hebbian and neuromodulatory mechanisms interact to trigger associative memory formation. Proceedings of the National Academy of Sciences, 111(51), E5584-E5592.

Johnson, H. A., Goel, A., & Buonomano, D. V. (2010). Neural dynamics of in vitro cortical networks reflects experienced temporal patterns. Nature neuroscience, 13(8), 917-919.

Kadowaki, T., & Nishimori, H. (1998). Quantum annealing in the transverse Ising model. Physical Review E, 58(5), 5355.

Kamioka, H., Maeda, E., Jimbo, Y., Robinson, H. P., & Kawana, A. (1996). Spontaneous periodic synchronized bursting during formation of mature patterns of connections in cortical cultures. Neuroscience letters, 206(2), 109-112.

Kapur, S., & Seeman, P. (2001). Does fast dissociation from the dopamine D2 receptor explain the action of atypical antipsychotics?: A new hypothesis. American Journal of Psychiatry. 158(3), 360–369.

Karakida, R., Okada, M., & Amari, S. I. (2016). Dynamical analysis of contrastive divergence learning: Restricted Boltzmann machines with Gaussian visible units. Neural Networks, 79, 78-87.

Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A., & Noguchi, J. (2010). Structural dynamics of dendritic spines in memory and cognition. Trends in neurosciences, 33(3), 121-129.

Kerr, R. R., Grayden, D. B., Thomas, D. A., Gilson, M., & Burkitt, A. N. (2014). Coexistence of reward and unsupervised learning during the operant conditioning of neural firing rates. PloS one, 9(1), e87123.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. Cognitive processing, 8(3), 159-166.

Korte, B., & Vygen, J. (2012). Combinatorial optimization (fifth edition). Berlin: Springer.

Laje, R., & Mindlin, G. B. (2002). Diversity within a birdsong. Physical review letters, 89(28), 288102.

Laje, R., & Buonomano, D. V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. Nature neuroscience, 16(7), 925-933.

Lappalainen, H., & Honkela, A. (2000). Bayesian non-linear independent component analysis by multi-layer perceptrons. In Advances in independent component analysis (pp. 93-121). Springer London.

Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8595-8598). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Lee, T. W., Girolami, M., Bell, A. J., & Sejnowski, T. J. (2000). A unifying information-theoretic framework for independent component analysis. Computers & Mathematics with Applications, 39(11), 1-21.

Le Feber, J., Stegenga, J., & Rutten, W. L. (2010). The effect of slow electrical stimuli to achieve learning in cultured networks of rat cortical neurons. PLoS One, 5(1), e8871.

Legenstein, R., Pecevski, D., & Maass, W. (2008). A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. PLoS Comput Biol, 4(10), e1000180.

Lesch, K. P., & Waider, J. (2012). Serotonin in the modulation of neural plasticity and networks: implications for neurodevelopmental disorders. Neuron, 76(1), 175-191.

Linsker, R. (1992). Local synaptic learning rules suffice to maximize mutual information in a linear network. Neural Computation, 4(5), 691-702.

Linsker, R. (1997). A local learning rule that enables information maximization for arbitrary input distributions. Neural Computation, 9(8), 1661-1665.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature, 484(7394), 381-385.

Lotter, W., Kreiman, G., & Cox, D. (2016). Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. arXiv preprint arXiv:1605.08104.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. Nature neuroscience, 14(2), 154-162.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature, 503(7474), 78-84.

Mansour-Robaey, S., Mechawar, N., Radja, F., Beaulieu, C., & Descarries, L. (1998). Quantified distribution of serotonin transporter and receptors during the postnatal development of the rat barrel field cortex. Developmental brain research, 107(1), 159-163.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science, 275(5297), 213-215.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nature Reviews Neuroscience, 5(10), 793-807.

Markram, H. (2006). The blue brain project. Nature Reviews Neuroscience, 7(2), 153-160.

Markram, H., Gerstner, W., & Sjöström, P. J. (2011). A history of spike-timing-dependent plasticity. Frontiers in Synaptic Neuroscience, 3: 4.

Matsuzaki, M., Honkura, N., Ellis-Davies, G. C., & Kasai, H. (2004). Structural basis of long-term potentiation in single dendritic spines. Nature, 429(6993), 761-766.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature, 485(7397), 233-236.

Mitchell, M. (1998). An introduction to genetic algorithms. MIT press.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. Trends in Cognitive Science, 16(1), 72–80.

Mukai, Y., Shiina, T., & Jimbo, Y. (2003). Continuous monitoring of developmental activity changes in cultured cortical networks. Electrical Engineering in Japan, 145(4), 28-37.

Müllner, F. E., Wierenga, C. J., & Bonhoeffer, T. (2015). Precision of inhibition: dendritic inhibition by individual gabaergic synapses on hippocampal pyramidal cells is confined in space and time. Neuron, 87(3), 576-589.

Nishiyama, M., Togashi, K., Aihara, T., & Hong, K. (2010). GABAergic activities control spike timing-and frequency-dependent long-term depression at hippocampal excitatory synapses. Frontiers in synaptic neuroscience, 2, 22.

Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., & Prochiantz, A. (1984). Magnesium gates glutamate-activated channels in mouse central neurones.

Oja, E. (1982). Simplified neuron model as a principal component analyzer. Journal of mathematical biology, 15(3), 267-273.

Oja, E. (1989). Neural networks, principal components, and subspaces. International journal of neural systems, 1(01), 61-68.

Oja, E. (1992). Principal components, minor components, and linear neural networks. Neural networks, 5(6), 927-935.

Olshausen, B. A. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583), 607-609.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1?. Vision research, 37(23), 3311-3325.

van Os, J., Kenis, G., & Rutten, B. P. (2010). The environment and schizophrenia. Nature, 468(7321), 203–212.

Paille, V., Fino, E., Du, K., Morera-Herreras, T., Perez, S., Kotaleski, J. H., & Venance, L. (2013). GABAergic circuits control spike-timing-dependent plasticity. The Journal of Neuroscience, 33(22), 9353-9363.

Pawlak, V., & Kerr, J. N. (2008). Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. The Journal of Neuroscience, 28(10), 2435-2446.

Pawlak, V., Wickens, J. R., Kirkwood, A., & Kerr, J. N. (2010). Timing is not everything: neuromodulation opens the STDP gate. Spike-timing dependent plasticity, 138.

Pehlevan, C., & Chklovskii, D. B. (2014). A hebbian/anti-hebbian network derived from online non-negative matrix factorization can cluster and discover sparse features. In 2014 48th Asilomar Conference on Signals, Systems and Computers (pp. 769-775). IEEE.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint

arXiv:1511.06434.

Ramirez, S., Liu, X., Lin, P. A., Suh, J., Pignatelli, M., Redondo, R. L., ... & Tonegawa, S. (2013). Creating a false memory in the hippocampus. Science, 341(6144), 387-391.

Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. Nature, 413(6851), 67-70.

Royer, S., & Paré, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature, 422(6931), 518-522.

Ruaro, M. E., Bonifazi, P., & Torre, V. (2005). Toward the neurocomputer: image processing and pattern recognition with neuronal cultures. Biomedical Engineering, IEEE Transactions on, 52(3), 371-383.

Salgado, H., Köhr, G., & Trevino, M. (2012). Noradrenergic 'tone'determines dichotomous control of cortical spike-timing-dependent plasticity. Scientific reports, 2.

Savin, C., Joshi, P., & Triesch, J. (2010). Independent component analysis in spiking neurons. PLoS Comput Biol, 6(4), e1000757.

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10(5), 1299-1319.

Schrödinger, E. (1992). What is life?: With mind and matter and autobiographical sketches. Cambridge University Press.

Schulz, J. M., Redgrave, P., & Reynolds, J. N. (2010). Cortico-striatal spike-timing dependent plasticity after activation of subcortical pathways. Frontiers in synaptic neuroscience, 2, 23.

Seol, G. H., Ziburkus, J., Huang, S., Song, L., Kim, I. T., Takamiya, K., ... & Kirkwood, A. (2007). Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. Neuron, 55(6), 919-929.

Shahaf, G., & Marom, S. (2001). Learning in networks of cortical neurons. The Journal of Neuroscience, 21(22), 8782-8788.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, & J. L. McClelland (Eds.), Parallel distributed processing (pp. 194–281). The MIT Press.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nature neuroscience, 3(9), 919-926.

Sumbre, G., Muto, A., Baier, H., & Poo, M. M. (2008). Entrained rhythmic activities of neuronal ensembles as perceptual memory of time interval. Nature, 456(7218), 102-106.

Sussillo, D., & Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron, 63(4), 544-557.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. Machine learning, 3(1), 9-44.

Takekawa, T., Isomura, Y., & Fukai, T. (2010). Accurate spike sorting for multi‐unit recordings. European Journal of Neuroscience, 31(2), 263-272.

Tetzlaff, C., Okujeni, S., Egert, U., Wörgötter, F., & Butz, M. (2010). Self-organized criticality in developing neuronal networks. PLoS Comput Biol, 6(12), e1001013.

Toyoizumi, T., Pfister, J. P., Aihara, K., & Gerstner, W. (2005). Generalized Bienenstock–Cooper–Munro rule for spiking neurons that maximizes information transmission. Proceedings of the National Academy of Sciences of the United States of America, 102(14), 5239-5244.

Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. Nature Reviews Neuroscience, 5(2), 97-107.

Urbanczik, R., & Senn, W. (2009). Reinforcement learning in populations of spiking neurons. Nature Neuroscience, 12(3), 250-252.

van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science, 274(5293), 1724.

Weingessel, A., & Hornik, K. (2000). Local PCA algorithms. IEEE Transactions on neural Networks, 11(6), 1242-1250.

Wu, S., Nakahara, H., & Amari, S. I. (2001). Population coding with correlation and an unfaithful model. Neural Computation, 13(4), 775- 797.

Xu, L. (1993). Least mean square error reconstruction principle for self-organizing neural-nets. Neural networks, 6(5), 627-648.

Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C., Urakubo, H., Ishii, S., & Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. Science, 345(6204), 1616-1620.

Zhang, J. C., Lau, P. M., & Bi, G. Q. (2009). Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. Proceedings of the National Academy of Sciences, 106(31), 13028-13033.

Zhou, Q., Homma, K. J., & Poo, M. M. (2004). Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. Neuron, 44(5), 749-757.

Achievements

Papers

[1] Tanaka Y, Isomura T, Shimba K, Kotani K, Jimbo Y. Neurogenesis enhances response specificity to spatial pattern stimulation in hippocampal cultures. IEEE Transactions on Biomedical Engineering, In press (2017).

[2] Isomura T, Sakai K, Kotani K, Jimbo Y. Linking Neuromodulated Spike-timing Dependent Plasticity with the Free-energy Principle. Neural Computation, 28(9): 1859–1888 (2016).

[3] Isomura T, Toyoizumi T. A Local Learning Rule for Independent Component Analysis. Scientific Reports, 6: 28073 (2016).

[4] Isomura T, Kotani K, Jimbo Y. Cultured Cortical Neurons Can Perform Blind Source Separation According to the Free-Energy Principle. PLoS Computational Biology, 11(12): e1004643 (2015).

[5] Isomura T, Shimba K, Takayama Y, Takeuchi A, Kotani K, Jimbo Y. Signal Transfer Within a Cultured Asymmetric Cortical Neuron Circuit. Journal of Neural Engineering, 12(6): 066023 (2015).

[6] Isomura T, Ogawa Y, Kotani K, Jimbo Y. Accurate Connection Strength Estimation Based on Variational Bayes for Detecting Synaptic Plasticity. Neural Computation, 27(4): 819–844 (2015).

[7] Shimba K, Sakai K, Isomura T, Kotani K, Jimbo Y. Axonal Conduction Slowing Induced by Spontaneous Bursting Activity in Cortical Neurons Cultured in a Microtunnel Device. Integrative Biology, 7(1): 64–72 (2015).

[8] 関本正志, 下野勇希, 赤尾旭彦, 磯村拓哉, 小川雄太郎, 黄琦, 小谷潔, 神保泰彦. 視覚と聴覚による多感覚拡張現実感技術を用いた Brain-Computer Interface の開発, 電気学会論文誌 C, 136(9): 1291–1297 (2016).

[9] 田中幸美, 門倉智之助, 磯村拓哉, 榛葉健太, 小谷潔, 神保泰彦. 培養海馬ニューロン新生の制御によるパターン分離能力の向上, 電気学会論文誌 C, 135(7): 805–812 (2015).

International conferences

[1] Isomura T, Toyoizumi T, Kotani K, Jimbo Y: Developing an in vitro system for investigating neuronal linear system identification; Ann Meet Soc Neurosci (SfN2016), 849.01, San Diego (USA), 16 November 2016.

[2] Isomura T, Kotani K, Jimbo Y. Maximum Entropy Learning in Cultured Cortical Neural Networks. Joint 8th Int Conf on Soft Comput and Int Sys and 17th Int Symp on Adv Int Sys (SCIS&ISIS2016), Sa5-2-(5) (#1570264712), Hokkaido (Japan), 27 August 2016.

[3] Isomura T, Kotani K, Jimbo Y. Functional Similarities Between Machine Learning and Cultured Neural Networks. 10th FENS Forum of Neuroscience, B050, Copenhagen (Denmark), 5 July 2016.

[4] Isomura T, Kotani K, Jimbo Y. Neuronal Maximum a Posteriori Estimation on Microelectrode Arrays. MEA Meeting 2016, doi: 10.3389/conf.fnins.2016.93. 00120, Reutlingen (Germany), 1 July 2016.

[5] Isomura T, Sakai K, Sato Y, Kotani K, Jimbo Y. An Information-Theoretical Interpretation for Neural Modulation of Spike-Timing Dependent Plasticity: Towards Cellular-Based Computational Psychiatry. Ann Meet Soc Neurosci (SfN2015), 94.14, Chicago (USA), 17 October 2015.

[5] Tanaka Y, Isomura T, Shimba K, Kotani K, Jimbo Y. Distance Dependent Activation of Dissociated Hippocampal Network by Tetanic Stimulation. 37th Ann Int Conf IEEE EMBS (EMBS2015), Milan (Italy), 28 August 2015.

[6] Isomura T, Kotani K, Jimbo Y. Blind source separation performed by dissociated culture of rat cortical neurons is consistent with the free-energy principle. Ann Meet Soc Neurosci (SfN2014), 188.21, Washington DC (USA), 16 November 2014.

[7] Isomura T, Kotani K, Jimbo Y. Cultured cortical neurons use the inverse recognition model for blind source separation. 9th FENS Forum of Neuroscience, G020, Milan (Italy), 7 July 2014.

[8] Isomura T, Kotani K, Jimbo Y. Cultured Cortical Neurons Can Separate Source Signals From Mixture Inputs. MEA Meeting 2014, A16, Reutlingen (Germany), 3 July 2014.

国内学会・シンポジウム等（筆頭著者としての場合のみ記載）

[1] 磯村拓哉, 小谷潔, 神保泰彦. 培養神経回路網の機械学習的な側面. 第 39 回日本神経科学大会, O1-H-5-4, 横浜, 2016 年 7 月 20 日.

[2] 磯村拓哉. Cultured Cortical Neurons Can Perform Blind Source Separation According to the Free-Energy Principle. Consciousness Club Tokyo (Organized by Dr. Ryota Kanai), 東京, 2016 年 3 月 31 日.

[3] 磯村拓哉, 小谷潔, 神保泰彦. 神経回路網に無意識的推論は宿るか？. 平成 27 年度電気学会 医用・生体工学研究会, MBE-16-27, 東京, 2016 年 3 月 22 日.

[4] 磯村拓哉. Friston の free energy principle: hierarchical model と dynamic expectation maximization の周辺について. 産総研 第 1 回大脳皮質モデル勉強会(一杉裕志先生主催), 東京, 2015 年 11 月 24 日.

[5] 磯村拓哉, 小谷潔, 神保泰彦. 統合失調症陽性症状の in silico/in vitro モデルの構築. 計測自動制御学会 ライフエンジニアリング部門シンポジウム (LE2015), SR-7(1L-11, 108-111), 福岡, 2015 年 9 月 2 日.

[6] 磯村拓哉, 小谷潔, 神保泰彦. 神経調節物質によるスパイク時刻依存可塑性修飾の理論. 第 54 回 日本生体医工学会大会, P3-2-28-B, 2015 年 5 月 7 日.

[7] 磯村拓哉, 小谷潔, 神保泰彦. 神経回路網における predictive coding に関する基礎的研究. 平成 26 年度電気学会 医用･生体工学研究会, MBE-15-34, 東京, 2015 年 3 月 27 日.

[8] 磯村拓哉, 小谷潔, 神保泰彦. ドーパミンは STDP とスパイキングニューロンが実行する PCA の精度を調節する, 第 37 回日本神経科学大会, P1-375, 横浜, 2014 年 9 月 11 日.

[9] 磯村拓哉, 小谷潔, 神保泰彦. 神経回路網における最適モデル選択――シミュレーションを用いた基礎的研究. 平成 26 年度 電気学会 C 部門大会, TC4-23, 島根, 2014 年 9 月 4 日.

受賞等

[1] SCIS&ISIS2016 Best Student Presentation Award, 2016 年 10 月受賞.

[2] 2016 年度 JNS-SfN Exchange Travel Award, 2016 年 8 月採択.

[3] FENS-IBRO/PERC Travel Grant 2016, 2016 年 3 月採択.

[4] 平成 27 年度 電気学会 C 部門技術委員会奨励賞, 一般社団法人 電気学会, 2016 年 3 月受賞.

[5] 平成 27 年度 技術交流助成 研修プログラム（海外研修）, 公益財団法人 中谷医工計測技術振興財団, 2016 年 2 月採択.

[6] 平成 27 年度 海外渡航奨学金, 公益財団法人 精密計測技術振興財団, 2015 年 9 月採択.

[7] 平成 27 年度 計測自動制御学会 生体・生理工学部会 学生奨励賞, 公益社団法人 計測自動制御学会, 2015 年 9 月受賞.

[8] 平成 26 年度 電気学会 C 部門大会 優秀発表賞, 一般社団法人 電気学会, 2015 年 8 月受賞.

[9] 平成 25 年度 第一種奨学金 学業優秀による全額免除, 独立行政法人 日本学生支援機構, 2014 年 5 月受賞.

Supplementary Information

Unconscious Inference in Neural Networks: Electrophysiology and Learning Theory

Takuya Isomura

Graduate School of Frontier Sciences, The University of Tokyo

Because the contents of Supplementary Information will be published from a scientific journal soon, they will not be published on the internet for five years.