Inverse Bayesian inference as a key of consciousness featuring a macroscopic quantum logical structure

Gunji, Yukio-Pegio
Shinohara, Shuji
Haruna, Taichi
Basios, Vasileios

(Citation)
Biosystems,152:44-65

(Issue Date)
2017-02

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90004099

Inverse Bayesian Inference as a Key of Consciousness Featuring a Macroscopic Quantum Logical Structure

Yukio-Pegio Gunji[1,*], Shuji Shinohara[2], Taichi Haruna[3], Vasileios Basios[4]

[1]Department of Intermedia Art and Science,
School of Fundamental Science and Technology, Waseda University, Ohkubo 3-4-1, Shinjuku-ku, Tokyo, 169-8555, Japan
[2]Graduate School of Medicine, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan
[3] Department of Planetology, Faculty of Science, Kobe University, Rokkod-dai 1-1, Nada, Kobe 657-8501
[4] Department of Statistical Physics and Complex Systems, Université Libre de Bruxelles, Boulevard du Triomphe, B-1050 Brussels, Belgium

*yukio@waseda.jp

Abstract

To overcome the dualism between mind and matter and to implement consciousness in science, a physical entity has to be embedded with a measurement process. Although quantum mechanics has been regarded as a candidate for implementing consciousness, nature at its macroscopic level is inconsistent with quantum mechanics. We propose a measurement-oriented inference system comprising Bayesian and inverse Bayesian inferences. While Bayesian inference contracts probability space, the newly defined inverse one relaxes the space. These two inferences allow an agent to make a decision corresponding to an immediate change in their environment. They generate a particular pattern of joint probability for data and hypotheses, comprising multiple diagonal and noisy matrices. This is expressed as a nondistributive orthomodular lattice equivalent to quantum logic. We also show that an orthomodular lattice can reveal information generated by inverse syllogism as well as the solutions to the frame and symbol-grounding problems. Our model is the first to connect macroscopic cognitive processes with the mathematical structure of quantum mechanics with no additional assumptions.

1. Introduction

Since Chalmers established that the issue of understanding consciousness and qualia is an extraordinarily difficult problem (Chalmers,1996), various researchers have approached it in different ways. Recent approaches based on phenomenal consciousness in neuroscience, robotics and philosophy have brought us closer to a possible solution, where the phenomenal consciousness could lack the nature of subjectivity relevant for measurement. While these approaches can be viewed as converging toward the dynamic nature of matter and quality, they need the measurement-oriented notion. This is consistent with endophysics (Rössler,1996; 1998, Atmanspacher et al., 2002, Atmanspacher, 2003) or internal measurement in science (Matsuno, 1989, Gunji, 1994, Gunji & Kusunoki, 1997) (which we refer to collectively as endo perspective) as well as neutral monism in philosophy (Silberstein & Chemero, 2015, Strawson, 2006). We propose a model of the measurement process based on the contraction and relaxation of its probability space to implement such a dynamical nature. The connection between consciousness, phenomenal consciousness, neuroscience theories, neutral monism, and endo perspective is not clear. As a result, we first clarify and establish these relations.

When Chalmers conferred hard-problem status on consciousness and qualia, an essential difference between matter and mind in nature became accepted, at which point his idea was classified as naturalistic dualism (Chalmers, 2007). A subjective quality cannot be reduced to a physical property, and vice versa. Consequently, many scientists evaded this issue since it seemed incapable of being solved in principle. They simply accepted naturalistic dualism in the same way as they had done based on previous studies (Popper & Eccles, 1977).

The failure of classical artificial intelligence (AI) could lead to a sense of which fragments of proto-intelligence or knowledge are not in a center of consciousness (i.e., a particular local area) but instead are embedded in environments (Pfeifer & Scheier, 2001). This leads to the idea of subsumption architecture (Brooks, 1986; 1991) and/or morphological computing (Pfeifer et al., 2007). In trying to compute how to bend metallic robotic fingers to pick up a raw egg without breaking it, classical AI fails because of the amount of computation required. If the robotic fingers are covered with a rubber skin that can adequately absorb physical shocks, the task can be achieved without an excessive amount of computation. The rubber-skin interface allows negotiation of the physical world in which the egg exists and the virtual world in which symbolic manipulation can be programmed. In that sense, the interface is a type of body. Hence, constructing the interface as a physical body is referred to as an

embodiment of intelligence (Varela et al., 1991). Although fragments of proto-intelligence are embedded in the body, the question arises as to whether intelligence exists in a programmable manipulation. If it does not, intelligence as a whole could be embedded not only in the body, but also in the environments surrounding it (Varela, 1997, Pfeifer & Gomez, 2009).

In subsumption architecture, a system of multi-agents plays a role in the interface. Each agent is merely a simple system following an equally simple rule, with no intrinsic intelligence. Contingent temporal configurations of agents are an embodiment of intelligence in a multi-agent system. Intelligence is not carried by a central system, but rather it arises collectively from the multi-agent system (Reynolds,1987, Couzin et al., 2002, Olfati-Saber, 2006).

These ideas are consistent with the notion of phenomenal consciousness (Tye,1997, Clark, 1998; 2003) developed in philosophy and cognitive science based on Husserl and Hedeggar's the phenomenology (Husserl, 1913=2001, Heideggar, 1927=1996). In phenomenology, anything is comprehensible by its surroundings. The relation between an object and its surroundings is embedded at each local site. When a particular function in the world appears as a concrete thing, such an object with a particular function is used as a particular tool ("presence-at-hand"); it is then bodily connected to the agent who uses it as a part of their body ("readiness-to-hand") (Heideggar, 1927=1996, Clark, 1998; 2003). A dynamic network of functional connections can give rise to a dynamic change of the owned body and/or consciousness as a whole; that is what is meant by phenomenal consciousness. The notion of a "thing" in the world can be extended to a human body, and the notion of usability can be extended to bodily sensations (Gallagher & Zahavi, 2008, Jaegher et al., 2010). A sense of bodily agency and/or ownership (Tsakiris et al., 2006, Synofzik et al., 2008) can also be comprehended in the framework of phenomenal consciousness (Gallagher, 2000).

How are matters in neuroscience? Koch, who focused on neural correlates of consciousness (NCC) (Rees et al., 2002), also recently abandoned mind/matter dualism and confessed his sympathy for panpsychism in which mind can be contained in anything (Koch, 2012). Tononi, who proposed information integration theory (IIT) (Tononi, 2008, Oizumi et la., 2014), also proposed an idea based on panpsychism (Balduzzi & Tononi, 2009, Tononi & Koch, 2016). However, their ideas are consistent with phenomenal consciousness rather than panpsychism. After the finding of readiness potential (Libet et al., 1983, Haggard et al., 2002, Frith et al., 2000), intentional consciousness is regarded as an area employed in postdiction (Koch, 2012, Maeno, 2005). Neural networks used in the readiness potential are referred to as unconscious zombies. After the activities of these unconscious zombies, a neural area correlated with the intentional consciousness can interpret a voluntary action

triggered not by the zombies but by the intentional consciousness itself (Koch, 2012, Maeno, 2005). Most neuroscientists, including Koch and Tononi, accept these views. A population of unconscious zombies can be compared to the system of multi-agents in subsumption architecture, to the layer of rubber skin in robotics, and to the dynamic network with respect to a functional connection. Although IIT is used to detect the intrinsic difference between an intent-wholeness (unity as a whole) and an extent-wholeness (sum of parts) (Tononi, 2008), it can also be considered as a way to estimate the relationship of intentional consciousness part and the population of zombies. Therefore, these theories are consistent with phenomenal consciousness.

Intentional consciousness covered by a population of zombies and body can sometimes include objects outside the body and exclude those inside it (Clark, 1998, 2003). That is an optimization process adapted to a given environment. Since both neuroscience and cognitive science focus on such an optimization, they tend to use only Bayesian inference in describing cognitive processes (Gigerenzer & Hoffrage, 1995, Knill & Pouget, 2004, Manktelow, 2012). Bayesian inference can reduce or contract the probability space dependent on empirical data, allowing the optimal solution to be found more readily. In contrast, the hypothesis of a global workspace (GWS) in neuroscience refers not only to a similar contraction of the probability space, but also to its expansion (Dehaene et al., 1998, Dehaene & Naccache, 2001, Dehanene & Changeux, 2011). A particular internally selected neural activity can be globally connected and propagated to various areas of the brain (Singer & Gray, 1995), which implies expansion of the space. The process of GWS might be directly related to generating a local singular structure, i.e., a “bundle” of qualities including qualia. We shall return to this issue later.

The question arises as to whether the singularity and/or locality carried in consciousness and qualia can be comprehended in phenomenal consciousness. In the case of rubber skin or the configuration of agents, many logical and programmable computational processes are not implemented directly in the system but could be indirectly embedded in non-logical material in a local area. In other words, the relation between logical states is embedded at a local site. If this embedding converges to a singular state under infinite recursive iteration, then this singular self-similar state (Scott, 1972, Gunji & Toyoda, 1997) is a candidate for local and singular states such as qualia and quality. However, the notion of phenomenal consciousness does not restrict the locality and singularity of consciousness and quality. A singular and local quality does not exist intrinsically, but appears instead as a phenomenon in the perspective of phenomenal consciousness. Although qualia and the subjective quality might be addressed by phenomenal consciousness, they could appear as illusions.

The proponents of singularity and locality of qualia and consciousness have moved to panqualityism or neutral monism (Strawson, 2006) because panpsychism has failed and suffers from a combinational problem (Chalmers, 2015). Panqualityism addresses the view that any physical thing can be endowed with quality, i.e., fragments of proto-qualia (Coleman, 2012). In mind/body dualism, there is the notion that if someone with a body temperature of 35°C touches a physical object whose temperature is 50°C, the sensation of hotness appears in the person's mind. In panqualityism, even this sensation is embedded in the physical object. While temperature is a quantity, hotness is a quality. Although there is an intrinsic difference between quantity and quality in nature, a 15°C temperature difference can cause the sensation of hotness. If this differential structure is recursively embedded in a local site in a self-similar manner, a singular state could arise whose quality is so simple that it reveals not a dynamic nature but a static one (Scott, 1972). We have previously attempted a similar but dynamic self-similar construction for dynamics (Gunji, 1994, Gunji et al., 1997, Gunji & Toyoda, 1997).

However, one final question arises. How is it possible to have a physical object embedded with a subjective quality? Put differently, how is neutral monism implemented in science? This may seem to be more akin to spirituality. Although there have been some previous attempts—the quantum consciousness hypothesis, in which quantum coherence occurs in cytoskeletal structures (Hameroff & Penrose, 1996a, b), and the idea that the water mass in the brain could generate consciousness (Jibu et al., 1994) - the criticism has been made that quantum effects do not contribute to macroscopic phenomena at normal temperatures (Grush & Churchland, 1995). While the internal quantum state in a molecule is shielded from thermal fluctuations and is considered in millikevin range (Matsuno & Paton, 2000, Igamberdiev & Shklovskiy-Kordi, 2016), the possibility of quantum brain theory is controversial.

There is only one way to ground such a perspective in science. A real thing that was considered as being independent of observation is now considered as a physical thing that has already been filtered by observation or measurement. Thus, the physical thing contains a measurement process in its base. Although it has been around 30 years since endoperspective was first proposed (Matsuno, 1989, Rössler, 1998), its significance has not yet been clarified. Such clarity will not be possible until science addresses the issue of consciousness (Strawson, 2006, Skrbina, 2009, Silberstein & Chemero, 2015, Seager, 2012).

The endoperspective is not sufficiently mature to comprehend the notion of consciousness. To address the neutral monism and to express a form of quality, non-local things have to be embedded at a local site, and local things have to be extended to non-local ones. This can lead to a local singular structure: a bundle of qualities containing temperature,

perception, feeling, and emotion (Strawson, 2006, Skrbina, 2009, Silberstein & Chemero, 2015, Seager, 2012). Dynamical interpolation between the local and non-local can influence the resulting perspective in cognition, local reductionism, and non-local non-reductionism. These complex cognitive natures are not addressed by the endoperspective. The question also arises as to whether or not the object that is accompanied by the measurement process is related to quantum mechanics. Independent of the quantum consciousness hypothesis, can the endoperspective lead to quantum mechanics (Svozil,1993)? Does the weak quantum mechanics proposed in the endoperspective contribute to the understanding of consciousness (Atmanspacher, 2003, Atmanspacher & Graben, 2015)?

In order to answer these questions, we propose a measurement model featuring an inference process. We focus on Bayesian inference for the contraction process of the probability space, and a newly proposed inverse Bayesian inference for its relaxation process. As mentioned before, this construction is entirely consistent with the hypothesis of GWS, so much so that making decisions based on Bayesian and Inverse Bayesian (BIB) inference can be regarded as a simple cognitive model. We first show that inverse Bayesian inference can cope with drastic changes in environments, while Bayesian inference can only optimize under a stable environment. Secondly, we show that BIB inference can generate the perspective of a pasted universe consisting of diagonal matrices of joint probabilities of data and hypotheses. Thirdly, we show that such a pasted universe consisting of diagonal matrices can be expressed as an orthomodular lattice that can correspond to quantum logic. That is the first model to connect the mathematical structure of quantum mechanics with a macroscopic cognitive process without the principle of complementarity.

2. Materials and Methods

2-1. BIB inference implementation

Basic Definition

No object exists without an observer; it is destined to be filtered by observation or measurement. An object described like this is referred to here as a measurement-oriented object (Gunji, 2004, Gunji & Kamiura, 2004). How can one express a measurement-oriented object? If the measurement process is expressed as a map f, and an object to be measured is represented by x, a measurement-oriented object might be expressed as $f(x)$. However, the measurement or observation process is not a well-defined map, but rather a one-to-many mapping, i.e., a map opened to impossible

alternatives (Matsuno, 1989, Gunji, 1994, Gunji et al., 1997, Gunji, 2004). This implies that a mapped image embedded with probability (i.e., possible alternatives) is inferred by a particular inference system, while the inference system is also perpetually revised and modified (i.e., impossible alternatives can be included). We define such a measurement-oriented object using BIB inference.

Although Bayesian inference has been used in cognitive and neuroscience (Gigerenzer & Hoffrage, 1995, Knill & Pouget, 2004, Manktelow, 2012), inverse Bayesian inference mentioned by Arecchi (Arecchi, 2003; 2011) has not been used (Gunji et al., 2016). Bayesian inference is a strategy for arriving immediately at the optimal solution. Depending on the data (i.e., an empirical condition), the probability of a particular hypothesis is changed. The probability of a hypothesis is temporally replaced by a conditional probability under the condition of taking data. Thus, the probability of a particular hypothesis fitting the given data grows increasingly.

In consequence, Bayesian inference *contracts* the distribution of probability because only hypotheses that fit the given data are taken into consideration in making a decision; any other hypothesis is ignored. The probability of an event is temporally replaced by the probability of only part of it. If only contraction were to occur, then a universe consisting of events would shrink to a point. A subject using only Bayesian inference is destined to reach a dead end. A different process is required that *relaxes* the probability to implement a measurement-oriented object. A contraction/relaxation pair would be more appropriate for an open, indefinite world.

A contraction/relaxation probability pair is a generalization of a cause–effect loop memory. A cause–effect relation can be expressed as an inclusion one. The statement “men implies invertebrates” expresses the notion that the set “men” is included in the set “invertebrates”. Thus, a relation in which a cause implies an effect is expressed as a pair of lines (one shorter than the other) in a cause–effect triangle, as shown in Fig. 1. The implication from a cause to its effect is expressed as a downward arrow. If a particular cause–effect relation is repeatedly perceived, then an upward arrow from an effect to a cause is also kept, i.e., a particular cause entailing a particular effect is expected. Thus, a cause–effect loop is kept whenever a cause–effect relation is empirically obtained and is expected or anticipated.

Fig. 1 also shows how a cone, not a triangle, is obtained in a cognitive process. This is illustrated in the experiment of choice blindness (Johansson et al., 2005). In that experiment, a male subject is shown photographs (right and left) of two different women, and is asked to state his preference. After the subject has made his choice, the experimenter passes him the apparently chosen photograph, but in fact has switched it with the other one without the

subject noticing. The subject is then asked to explain his choice, usually doing so while being unaware that he is actually discussing the rejected photograph. This is why the phenomenon is referred to as choice blindness.

A pair of cause–effect loops corresponds to the previous and invented cause–effect processes. If we suppose that he actually chose the left photograph, then the left cause–effect loop was being kept initially. However, he invented the right cause–effect loop after the photograph was replaced. The upper black circle represents a current event. Because the subject observed the right photograph while explaining his reason, a current event is facing it. By contrast, no event is now connected to the left cause–effect loop. The upper white circle represents the absence of a current event. In fact, “current” is a time interval. Thus, multiple cause–effect loops are bundled to construct a cause–effect cone in which the current event is integrated with multiple possible events. A gray current event represents the integration of black and white current events (Fig. 1).

Multiple cause–effect loops are integrated. The inclusion relation of a cause–effect one is ill-defined and is expressed as a form of interpolating system. A particular cause is located in a particular effect as a context. At that time, the context could be revised and modified to embed the cause. Imagine a relationship between a local site of a city and the entire city. If you are lost in a particular place, you are aware of where you are if you can identify the particular place in the city. This implies you made a context (whole city) to embed a particular place. By contrast, when you choose a particular representative place in the city, the meaning of the city is revised depending on your choice of representative. The significance or meaning of a place and the city is temporally changed and revised because of the interpolation. This results in an alternation between contraction and relaxation of the meaning of the city. From the local to the global (e.g., awareness of lost place), the context of the global is expanded and revised; from the global to the local (e.g., choice of the representative), the context is chosen and contracted. That is also the case for a cause–effect loop.

If the cause–effect loop cone is implemented in the framework of probability, then both the probability and conditional probability of an event can correspond to a cause and an effect in a loop. Thus, we can generalize the contraction and relaxation processes in the framework of probability, i.e., a pair of Bayes and Inverse Bayes (BIB) inferences (Fig. 1, left). As mentioned before, Arecchi first proposed inverse Bayesian inference (Arecchi, 2003; 2011), and we subsequently modified his ideas (Gunji et al., 2016). In the present paper, we naturally expand our modification to define inverse Bayesian inference anew in a symmetric manner to Bayesian inference. Arecchi used Bayes’ formula and termed a particular usage of it as inverse Bayesian inference. Given Bayes’ formula, $P(d|h)P(h) = P(h|d)P(d)$, if one obtains the posterior probability of $P(h)$ as $P(h|d) = P(d|h)P(h)/P(d)$, this usage of Bayes’ formula is

called Bayesian inference. If one obtains the prior probability as $P(h) = P(h|d)P(d)/P(d|h)$, Arecchi calls this usage of Bayes' formula the inverse Bayesian inference (Arecchi, 2003; 2011). However, the essence of Bayesian inference is the replacement of $P(h)$ with $P(h|d)$. Thus, we naturally expand this idea inversely, and define inverse Bayesian inference by replacing $P(d|h)$ with $P(d)$, different from Arechi. That is why BIB inference can correspond to both the contraction and relaxation of the probability.

In neuroscience, a part of the global workspace theory (Dehanene et al., 1998, Dehaene & Naccache, 2001, Dehanene & Changeux, 2011) can be implemented by Bayesian inference (Arecchi, 2003; 2011). Given a set of external stimuli, neural activities are locally synchronized. A local domain consisting of synchronized neurons can correspond to a hypothesis by which external stimuli (data) can be interpreted. There are multiple domains corresponding to multiple hypotheses, and the domain that has the largest synchronized domain is chosen. This could correspond to a Bayesian inference in which a hypothesis with the most probability is chosen (Arecchi, 2003; 2011). The global workspace theory, however, is not completed by Bayesian inference alone. The chosen hypothesis (i.e., a particular neural activity) is globally connected to all other areas of the brain, and the interpretation of a given set of data that is obtained by the chosen hypothesis can be accessed and used by any other area of the brain (Dehanene et al., 1998, Dehaene & Naccache, 2001, Dehanene & Changeux, 2011). That is the essential nature of global workspace theory. This property, openness to other areas, can be expressed as a relaxation process, and by replacing the probability of a part of the set by the probability of a set as a whole. In other words, the probability of a particular conditional event is replaced by the probability of general events. This process is the same as a relaxation one.

The anterior half of global workspace theory until an optimal neural activity is chosen can be expressed as the contraction of information, since external stimuli are collected to give rise to a particular local neural activity. The posterior half of the theory can be expressed as a relation of information, since a local neural activity is globally propagated to multiple areas of the brain. A contraction/relaxation informational pair is found in global workspace theory (Dehanene et al., 1998, Dehaene & Naccache, 2001, Dehanene & Changeux, 2011).

We now formalize BIB inference. A data set and a hypothesis set are expressed by $D = \{d_1, d_2, \ldots, d_n\}$ and $H = \{h_0, h_1, \ldots, h_m\}$, respectively. Because of the relation between conditional and joint probability, Bayes' formula is expressed as

$$P^t(h|d) = P^t(d|h)P^t(h)/(\sum_k P^t(d|h_k)\, P^t(h_k)), \tag{1}$$

where d and h represents an element in the set of data and hypotheses, respectively. Since

$P^t(h_k)$ represents the probability of a hypothesis at t-th step, h_k and initially $P^0(h_k)$ is homogeneously given (i.e., $P^0(h_0) = P^0(h_1) = \ldots = P^0(h_m) = 1/(m+1)$), $P^t(d|h_k)$ which is the probability of particular data, d, under a hypothesis, h_k (i.e., the likelihood of h_k) is also defined, $P^t(h|d)$ in the form of (1) can be obtained without empirical data. Here, Bayesian inference is expressed as

$$P^{t+1}(h) = P^t(h|d). \tag{2}$$

If only Bayesian inference is introduced, the likelihood of hypotheses is not changed, such as

$$P^{t+1}(d|h) = P^t(d|h). \tag{3}$$

For example, let H be $\{h_o, h_1\}$, and D be {0, 1}, where 0 and 1 represents the head and tail of the coin. Imagine that you have to determine the probability of coin toss for a special coin whose probability of the coin toss is not uniformly random. You have two hypotheses for the probability such that $P^0(0|h_0) = 1/3$ and $P^0(0|h_1) = 3/4$. Thus, $P^0(1|h_0) = 2/3$ and $P^0(1|h_1) = 1/4$. Initially $P^0(h_0) = P^0(h_1) = 1/2$. Under this setting, you have the first result of coin toss as 0. By using eq-(1), $P^0(h_0|0) = P^0(0|h_0)P^0(h_0)/(P^0(0|h_0)P^0(h_0) + P^0(0|h_1)P^0(h_1)) = (1/3)(1/2) / ((1/3)(1/2) + (3/4)(1/2)) = 4/13$, and $P^0(h_1|0) = P^0(0|h_1)P^0(h_1)/(P^0(0|h_0)P^0(h_0) + P^0(0|h_1)P^0(h_1)) = 9/13$. Due to eq-(2), $P^1(h_0) = P^0(h_0|0) = 4/13$ and $P^1(h_1) = P^0(h_1|0) = 9/13$. Since you obtain data of 0, the probability of hypothesis h_1 increases from 1/2 to 9/13, in which the probability of 0 is bigger than h_0. If the second result of coin toss is also obtained as 0, $P^2(h_1) > P^1(h_1)$ because $P^2(h_1) = P^1(h_1|0) = 81/97$.

Inverse Bayesian inference is implemented symmetrically to (2), and is expressed as

$$P^{t+1}(d|h_s) = P^t(d). \tag{4}$$

In eq-(4), the conditional probability, $P^{t+1}(d|h_s)$, under a particular hypothesis condition is replaced by empirical data (i.e., a given time series of data), $P^t(d)$. Because $P^{t+1}(d|h_s)$ represents the probability of a particular data occurrence in the hypothesis h_s, replacing this probability implies replacing the hypothesis itself. Here, $P^t(d)$ is defined by a normalized frequency of data in a particular time interval, M. Given a series of data such that

$$e^{t-M}, e^{t-M+1}, \ldots, e^t, \tag{5}$$

where e^w is an element of $\{d_1, d_2, \ldots, d_n\}$, if the number of occurrences of d in a series of (4) is represented by $f(d)$, then

$$P^t(d) = f(d)/M. \tag{6}$$

The interval M can be determined by saturation of the mutual information for a time series with a particular interval M. The problem arises as to how a particular hypothesis, h_s, in (3) is chosen. We introduce the least optimal hypothesis. Thus, h_s in (3) satisfies the condition that

$$\forall \ h' \in \{h_0, h_1, \ldots, h_m\}, \ P^t(h_s) \leq P^t(h'), \tag{7}$$

where the least optimal choice is affected by choice error. It implies that the time scale of inverse Bayes inference is much longer than that of Bayes inference. In practice, we calculate

$$Q(k) = \sum_{j=0}^{k}(1- P^t(h_j)), \tag{8}$$

and a real number, r, is chosen randomly with uniform distribution in the interval $[0, Q(m)]$. The least optimal hypothesis with choice error, h_s, is given by

$$Q(s\text{-}1)<r\leq Q(s). \tag{9}$$

By this scheme one hypothesis h_s is chosen and (4) is applied, on one hand. For any other hypotheses such that $g \in \{h_0, h_1, \ldots, h_m\}$ and $g \neq h_s$, the likelihoods of the hypotheses are not changed such that

$$P^{t+1}(d|g) = P^t(d|g). \tag{10}$$

In this scheme, the probability of a hypothesis is continually replaced by its conditional probability under a particular condition of data because of Bayesian inference. The conditional probability of data under a particular hypothesis representing the hypothesis itself is replaced by the empirical probability of the data because of inverse Bayesian inference. At any time step, both types of inference are applied to a set of probability.

Recall $H = \{h_o, h_1\}$, $D = \{0, 1\}$ $P^0(0|h_0) = 1/3$ and $P^0(0|h_1) = 3/4$ for the coin toss case. Initially $P^0(h_0) = P^0(h_1) = 1/2$. As mentioned before, $P^0(h_0|0) = 4/13$, and $P^0(h_1|0) = 9/13$ for the 0 coin case. Due to eq-(2), $P^1(h_0) = P^0(h_0|0) = 4/13$ and $P^1(h_1) = P^0(h_1|0) = 9/13$. Imagine

that you have a time series of coin toss, 1, 1, 0, 1, 1, 1. Thus the empirical data, $P^0(0) = 1/6$. Due to eq-(4) (i.e., inverse Bayesian inference), and $P^0(h_0) = P^0(h_1)$, the least optimal hypothesis is randomly chosen, for example, h_1, and the likelihood of h_1 is replaced by the empirical data. Thus you obtain $P^1(0|h_1) = 1/6$ and $P^1(0|h_0) = P^0(0|h_0) = 1/3$. If the second result of the coin toss is also 0, $P^2(h_1|0) = P^1(0|h_1)P^1(h_1)/(P^1(0|h_0)P^1(h_0) + P^1(0|h_1)P^1(h_1)) = (1/6)(9/13)/((1/6)(9/13)+(1/3)(4/13)) = 9/17$. $P^2(h_1) = P^1(h_1|0) = 9/17$ is much smaller than 81/97 for the only Bayesian inference. The difference results from the change of the likelihood of h_1, and i.e., inverse Bayesian inference.

We now consider the role of BIB inference with respect to the prediction (anticipation) and postdiction. Since the conditional probability $P(h|d)$ is expressed as $P(d, h)/P(d)$, we obtain

$$P^t(d, h) = P^t(h|d)P^t(d). \tag{11}$$

Substituting Bayesian inference, $P^{t+1}(h) = P^t(h|d)$, for (11), we obtain

$$P^t(d, h) = P^{t+1}(h)P^t(d). \tag{12}$$

It is easy to see that the joint probability $P^t(d, h)$ is approximated by the product of the probability of hypothesis in future and that of data at present, which implies that present data could be independent of anticipated hypothesis. Similarly, $P(d|h)$ is expressed as $P(d, h)/P(h)$, and then

$$P^t(d, h) = P^t(d|h)P^t(h). \tag{13}$$

Substituting inverse Bayesian, $P^{t+1}(d|h) = P^t(d)$ (i.e., $P^t(d|h) = P^{t-1}(d)$), for (13), we obtain

$$P^t(d, h) = P^t(h)P^{t-1}(d). \tag{14}$$

This also implies that two events, data in the past and hypothesis at present, could be independent of each other, where the form of the time delay in (14) is different from that in (12). Note that (12) and (14) are not actual transition rules. If only Bayesian inference is used, $P^t(h|d)$ is calculated, $P^{t+1}(h)$ is updated by $P^t(h|d)$, and then $P^{t+1}(d|h)$ becomes the same as $P^t(d|h)$. If BIB (Bayesian and inverse Bayesian) inference is used, $P^t(h|d)$ is calculated, $P^{t+1}(h)$ is updated by $P^t(h|d)$, and then $P^{t+1}(d|h_s)$ for the least optimal hypothesis, h_s, is also updated by $P^t(d)$ where $P^{t+1}(d|h)$ for any other hypotheses but h_s becomes the same as $P^t(d|h)$.

Equations (12) and (14) shows the significance of Bayesian and of inverse Bayesian inference, respectively. Since BIB inference uses both Bayesian and inverse Bayesian inference, both (12) and (14) are implemented in the inference. It implies that joint probability $P^t(d, h)$ is initially influenced by anticipated hypothesis and then influenced by postdiction of data.

The existence of pre- and postdiction in the independence assumption is the key difference between the two forms of inference. In Bayesian inference, independence between data and hypothesis is achieved by the predicted probability of the hypothesis, $P^{t+1}(h)$ in (12). This implies that the probability of the hypothesis is modified *a priori* to achieve the optimal solution immediately. This could correspond to contraction toward the observer's own optimal goal. By contrast, in inverse Bayesian inference, the independence is achieved by the postdiction probability of data, $P^{t-1}(d)$ in (14). This implies that the probability of data (an empirical "thing") is negotiated *a posteriori* to relax the over-contracted world.

Idealized Implementation of BIB inference

We now implement BIB inference under a particular idealization, where a contraction of the probability (i.e., Bayesian inference) results from a relaxation of the probability (i.e., inverse Bayesian inference), and the data and hypothesis spaces are symmetrical to each other due to Bayes' formula. We assume here that data and hypothesis sets are expressed as $\{d_1, d_2, \ldots, d_n\}$ and $\{h_1, h_2, \ldots, h_n\}$, respectively. Since the conditional probability is defined by $P^t(d|h)=P^t(d, h)/P^t(h)$, we obtain (11), $P^t(d, h) = P^t(h|d)P^t(d)$, and $P^t(d)$ is expressed as a summation of any $P^t(d, h_j)$. Thus,

$$P^t(d, h) = P^t(h|d) \sum_{j=0}^{m} P^t(d, h_j). \tag{15}$$

We now introduce the operation of relaxation. First, some hypotheses, h, are collected that satisfy the statement such that

$$\theta' \leq P(h) < \theta \text{ with } [0.0, 1.0]. \tag{16}$$

Hypotheses with the condition (16) (w hypotheses) are collected such as

$$h_{s(1)}, h_{s(1)}, \ldots, h_{s(w)}. \tag{17}$$

It is assumed that these hypotheses could constitute a whole hypothesis space for particular data d, and then the joint probability of data and hypotheses with respect to all hypotheses satisfying the condition is expressed as

$$\sum_{j=1}^{w} P^t(d, h_{s(j)}) = P^t(d). \tag{18}$$

It is also assumed that a summation of joint probability with respect to all hypotheses could constitute a summation of probability of all data, which is a universal set, and then

$$\sum_{j=0}^{m} P^t(d, h_j) = 1.0. \tag{19}$$

Equations (18) and (19) imply relaxation because a part of the probability is expanded and is regarded as the probability of a universal set. In this sense, (17) is replaced by

$$P^{t+1}(d, h_{s(j)}) = P^t(h_{s(j)}|d). \tag{20}$$

Since $P^t(h_{s(j)}|d) = P^t(d, h_{s(j)})/P^t(d)$, (18) and (19), for any i in {1, 2, .., w}, we obtain

$$P^{t+1}(d, h_{s(i)}) = P^t(d, h_{s(i)})) / \sum_{j=1}^{w} P^t(d, h_{s(j)}). \tag{21}$$

In the form of (21), both contraction and relaxation are embedded. The assumption of (18) reveals that a collection of particular hypotheses, $\{h_{s(1)}, h_{s(2)}, \ldots, h_{s(w)}\}$, is replaced by a whole set of hypotheses, and is expanded by the probability of a universal set (i.e., 1.0). Thus, it implies relaxation (inverse Bayesian inference), which could entail contraction (Bayesian inference) in the form of (20), in which a joint probability is replaced by a conditional one.

Since Bayes' formula could result in a symmetrical form such that

$$P^t(d, h) = P^t(d|h) \sum_{j=0}^{n} P^t(h, d_j), \tag{22}$$

one has to collect a particular subset of $d_{r(1)}, d_{r(2)}, \ldots, d_{r(w)}$ symmetrically, and assume that

$$\sum_{j=1}^{w} P^t(d_{r(j)}, h) = P^t(h). \tag{23}$$

As well as (19), it is assumed that summation of $P^t(d_j, h)$ for all j in {0,1, …, n} equals 1.0, which results in

$$P^{t+1}(d_{r(j)}, h) = P^t(d_{r(j)}|h) = P^t(d_{r(j)}, h) / \sum_{j=1}^{w} P^t(d_{r(j)}, h). \tag{24}$$

Since $P^{t+1}(d, h) = P^{t+1}(h, d)$, symmetrical Bayesian inference, which is expressed as (21) and (24), is naturally introduced by assuming inverse Bayesian inference, i.e., (18) and (23). These procedures are illustrated schematically in Fig. 2.

After the application of (21) to the distribution of the joint probability, (24) is similarly applied. Thus, we redefine Bayesian inference reduced from inverse Bayesian inference by

$$P^{t+\Delta t(1)}(d, h_{s(i)}) = P^{t+\Delta t(0)}(d, h_{s(i)}) / \sum_{j=p}^{q} P^{t+\Delta t(0)}(d, h_{s(j)}), \tag{25}$$

$$P^{t+\Delta t(2)}(d_{r(j)}, h) = P^{t+\Delta t(1)}(d_{r(j)}, h) / \sum_{j=p}^{q} P^{t+\Delta t(1)}(d_{r(j)}, h), \tag{26}$$

where $\Delta t(0)<\Delta t(1)<\Delta t(2)$ and for j=0, 1, …, w (i.e., p=0, q=w), $\Delta t(0)$=0, and $\Delta t(2)$=1/2, and for j=w+1, w+2, …, n (i.e., p=w+1, q=n), $\Delta t(0)$=1/2, and $\Delta t(2)$=1. After that, the following is applied to the distribution of the joint probability to evoke the effect of Bayesian inference, such as

$$P^{t+1}(d, h) = (P^{t+1}(d, h))^2. \tag{27}$$

This process implies enhancement of the effect of Bayes and inverse Bayes inference. By this recipe, the coupling of BIB inference is simply implanted by the choice of hypotheses and replacement of joint probability with conditional probability. As mentioned in later sections, this recipe can be extended to a partition of a set of hypotheses such as {$h_{s(1)}$, $h_{s(2)}$, …, $h_{s(w)}$}, {$h_{s(w+1)}$, $h_{s(w+2)}$, …, $h_{s(v)}$}, …, {$h_{s(u+1)}$, $h_{s(u+2)}$, …, $h_{s(m)}$}.

Given a distribution of joint probability $P^0(d_1, h_1)$, $P^0(d_1, h_2)$, …, $P^0(d_1, h_6)$, $P^0(d_2, h_1)$, $P^0(d_2, h_2)$, …, $P^0(d_2, h_6)$, $P^0(d_6, h_1)$, $P^0(d_6, h_2)$, …, $P^0(d_6, h_6)$, where H is divided into {h_1, h_2, h_3} and {h_4, h_5, h_6} and symmetrically D is divided into {d_1, d_2, d_3} and {d_4, d_5, d_6}. In this situation, for any k = 1, 2, …, 6 and for s = 1, 2, 3, $P^{\Delta t(1)}(d_k, h_s) = P^0(d_k, h_s) / (P^0(d_k, h_1)+P^0(d_k,$

$h_2)+P^0(d_k,\ h_3))$ and $P^{1/2}(d_s,\ h_k) = P^{\Delta t(1)}(d_s,\ h_k) \ /\ (P^{\Delta t(1)}(d_1,\ h_k)+P^{\Delta t(1)}(d_2,\ h_k)+P^{\Delta t(1)}(d_3,\ h_k))$. Similarly, for any k = 1, 2, …, 6 and for s = 4, 5, 6, $P^{1/2+\Delta t(1)}(d_k,\ h_s) = P^{1/2}(d_k,\ h_s) \ /\ (P^{1/2}(d_k,\ h_4)+P^{1/2}(d_k,\ h_5)+P^{1/2}(d_k,\ h_6))$ and $P^1(d_s,\ h_k) = P^{1/2+\Delta t(1)}(d_s,\ h_k) \ /\ (P^{1/2+\Delta t(1)}(d_4,\ h_k)+P^{1/2+\Delta t(1)}(d_5,\ h_k)+P^{1/2+\Delta t(1)}(d_6,\ h_k))$. Asynchronously joint probabilities are replaced by conditional probabilities by these manners. These operations are iterated through time.

Results

The significance of inverse Bayesian inference

First, a simple case study is calculated for the basic definition of BIB inference, (1)–(14). Sets of data and of hypotheses are defined by {0, 1} and $\{h_0, h_1, \ldots, h_m\}$, respectively, where

$$P^t(1|h_k) = k/L, \tag{28}$$

$$L = m+1. \tag{29}$$

One can think of a hypothesis, h_k, as a virtual bag containing k red balls represented by 1 and (L-k) white balls represented by 0. Given a time series of data, either 0 or 1, one can infer the probability of a hypothesis (i.e., a distribution of $P(h)$) by both forms of inference.

If the probability of data is invariant and stable, which means

$$\lim_{M\to\infty} f(d)/M = \text{const.}, \tag{30}$$

then the probability of a hypothesis obtained only by Bayesian inference is that same as that obtained by BIB inference, as shown in Fig. 3 where the results of the latter are slightly unstable due to the temporal replacement of the conditional probability, $P^{t+1}(d|h_s)$, by the empirical data (i.e., (4)–(6). In the simulation, L=10, M=30, and a time series of data is generated for $P(1)$=0.45. The smaller M is, the more sensitive the replacement of $P^{t+1}(d|h_s)$ with $P^t(d)=f(d)/M$ is. If M>20, the BIB can show similar behavior.

Now we compare the probability of a hypothesis obtained only by Bayesian inference with that obtained by BIB inference for a series of data that is suddenly changed at a particular time step. Fig. 4 shows an example of such a case. The vertical axis represents the conditional probability, $P^t(1|h_{\text{opt}})$, where h_{opt} is the optimal hypothesis with respect to the

probability, such that

$$\text{For any } h \in \{h_0, h_1, \ldots, h_m\},\ P^t(h) \leq P^t(h_{\text{opt}}). \tag{31}$$

Given a time series of data in which the probability of obtaining the datum “1” is 0.8 for the first 500 steps and 0.2 after that, Bayesian inference (green curve) cannot follow the sudden change in the probability of the given data (red curve), and instead traces the accumulated probability of the given data (blue curve). Since the hypotheses themselves, $P(1|h_0)$, $P(1|h_1)$, …, $P(1|h_m)$, are invariant with time, it is only Bayesian inference that can trace a time series of data by switching the optimal hypothesis. Thus, the trajectory of $P^t(1|h_{\text{opt}})$ is that of a step function (Fig. 4A).

Fig. 4B shows a trajectory of $P^t(1|h_{\text{opt}})$ obtained by BIB inference, where the time series of data is the same as that in Fig. 4A. It is clear to see that this type of inference (green curve) can trace the sudden change in the probability of the given data (red curve), not the accumulated probability, where there is a delay to trace it in a term of time interval of collecting data. The second feature of BIB inference is its stable trajectory. As shown in Figs. 4A and 4B, as long as the probability of the given data does not change, the trajectory of the inference also does not change and remains relatively stable. To investigate the reason for this, another implementation of inverse Bayesian inference is introduced and is compared to the original implementation.

Instead of the choice of the least optimal hypothesis, a random choice and one involving the most optimal hypothesis are introduced. For the random choice, instead of (6) (or (7, 8)), h_s in (3) (or h_s in (8)) is randomly chosen from $\{h_0, h_1, \ldots, h_m\}$. For the choice of the most optimal hypothesis, h_s satisfies the condition that

$$\forall\ h \in \{h_0,\ h_1,\ \ldots,\ h_m\},\ P^t(h) \leq P^t(h_s). \tag{32}$$

Both types of inference with h_s randomly chosen are shown in Fig. 5A. That with the optimal h_s is shown in Fig. 5B, where a given time series of empirical data is the same as those for the case of Fig. 4, i.e., 0.8 before step 500 and 0.2 after it. As well as the BIB inference with the least optimal h_s, both inferences can trace the sudden change in the time series of empirical data although they are unstable, as compared with the inference in the least optimal case. There is no qualitative difference between the random choice and the most optimal choice with respect the joint probability.

To spell out the significance of choosing the least optimal hypothesis in inverse Bayesian inference, we compared the inverse inferences based on the least optimal

hypothesis choice and the random choice with respect to the changeability of the hypothesis. Figure 6 shows the probability of a hypothesis and the conditional probability of obtaining datum “1” under each hypothesis obtained by BIB inference, where the inverse Bayesian inference is based on the least optimal hypothesis choice, and where the given probability of obtaining datum “1” is 0.6 before step 500 and 0.4 after.

Fig. 6A shows a time series of the probability of all hypotheses. At step 300, the hypothesis h_6 is the optimal one, and after step ~550 the optimal one is replaced by h_7, which is stable up to step 1000. This implies that the conditional probability of datum “1” under the optimal hypothesis is that under h_6 between steps 300 and 550, and that under h_7 between steps 550 and 1000. Figure 6B shows a time series of conditional probability of obtaining “1” under hypotheses h_6 and h_7. In taking a time series of the probability, it is easy to see that the conditional probability of the optimal hypothesis is relatively flat between steps 300 and 550, and that it is ~0.4 between steps 550 and 1000 (Fig. 6D). Figure 6C shows a time series of conditional probability of obtaining “1” under all hypotheses. Given that the least optimal hypothesis is perpetually replaced by the empirical data, and that most hypotheses except for h_6 and h_7 could be the least optimal after step 300, most hypotheses are perpetually replaced by the empirical data. Therefore, while most hypotheses are being replaced by others continually, an optimal hypothesis such as h_7 is not replaced but is maintained (Fig. 6C). That is why the inference resulting from the optimal hypothesis is stable in replacing the least optimal hypothesis by the empirical data.

Compared to the least optimal choice, the random choice in inverse Bayesian inference could give rise to unstable inference, since even the optimal hypothesis is continually being replaced by empirical data. Fig. 7A shows a time series of the probability of all hypotheses. Between steps 50 and 300, the hypothesis h_6 is the optimal one, and between steps 300 and 900 the optimal one is replaced by h_7. Fig. 7B shows a time series of conditional probability of obtaining “1” under hypotheses h_6 and h_7. Notwithstanding that hypotheses h_6 and h_7 are the optimal ones, they are replaced so often that the conditional probability of datum “1” under h_6 or h_7 is not invariant and is continually being changed. Therefore, the conditional probability of datum “1” under the optimal hypothesis mainly consists of h_6 and h_7, and is unstable as shown in Fig. 7D. Figure 7C shows a time series of the conditional probability of datum “1” under all hypotheses. There is no stable and invariant conditional probability under any hypothesis. Thus, the least optimal choice of a hypothesis in inverse Bayesian inference could contribute to the stable inference, although it is sensitive to the temporally sudden change in the empirical data.

Next, we describe the simulation results of BIB inference in the idealized implementation. Figure 8 shows the distribution of the joint probability of 20 data items and

20 hypotheses, which is developed by (25)–(27) given an initial condition of a random distribution of the joint probability. Here, the hypotheses are divided into $\{h_{s(1)}, h_{s(2)}, \ldots, h_{s(10)}\}$ and $\{h_{s(11)}, h_{s(12)}, \ldots, h_{s(20)}\}$, and the data into $\{d_{r(1)}, d_{r(2)}, \ldots, d_{r(10)}\}$ and $\{d_{r(11)}, d_{r(12)}, \ldots, d_{r(20)}\}$, respectively. Each set is assumed to be the larger set, which implies (18), (19), and (23). In Fig. 8, hypotheses and data are arranged in the order of $s(1)$, $s(2)$, …, $s(20)$ and $r(1)$, $r(2)$, …, $r(20)$. In the steady state at t=10, the distribution of the joint probability is articulated into a diagonal matrix area and homogeneous noisy area. The areas of $[s(1), s(10)]\times[r(1), r(10)]$ and $[s(11), s(20)]\times[r(11), r(20)]$ are diagonal matrix areas, and $[s(11), s(20)]\times[r(1), r(10)]$ and $[s(1), s(10)]\times[r(11), r(20)]$ are homogeneous noisy areas. In a diagonal matrix area, there is a one-to-one correspondence between data and hypotheses. This means that there is a unique hypothesis, h, for each datum, d, with a high joint probability of $P(d, h)$, and that any other joint probabilities are very low, i.e., effectively zero. Thus, one can arrange hypotheses and data to locate all high joint probabilities at the diagonal line. That is why it is called a diagonal matrix area. A noisy area consists of the low but >0 joint probabilities.

The partition of hypotheses or data that is assumed to be the larger set in the form of (18), (19), and (23) (i.e., inverse Bayesian inference) can be generalized for multiple partitions, such as $(S(1, 1), \ldots, S(1, w_1))$, $(S(2, 1), \ldots, S(2, w_2))$, …, $(S(q, 1), \ldots, S(q, w_q))$, and $(R(1, 1), \ldots, R(1, w_1))$, $(R(2, 1), \ldots, R(2, w_2))$,…, $(R(q, 1), \ldots, R(q, w_q))$, where for each $S(k, j)$, there exists a hypothesis h in $\{h_1, h_2, \ldots, h_n\}$ with one-to-one correspondence (for $R(k, j)$, d in $\{d_1, d_2, \ldots, d_n\}$ with one-to-one correspondence, respectively). For this partition, Bayesian inference reduced from inverse Bayesian inference is expressed as

$$P^{t+\Delta T(k,\,1)}(d, h_{S(k,\,j)}) = P^{t+\Delta T(k,\,0)}(d, h_{S(k,\,j)}) \,/\, \sum_{j=p}^{u} P^{t+\Delta T(k,\,0)}(d, h_{S(k,\,j)}), \tag{33}$$

$$P^{t+\Delta T(k,\,2)}(d_{R(k,\,j)}, h) = P^{t+T(k,\,1)}(d_{R(k,\,j)}, h) \,/\, \sum_{j=p}^{u} P^{t+\Delta T(k,\,1)}(d_{R(k,\,j)}, h), \tag{34}$$

where k=1, 2, …, q, $\Delta T(k, 0)=(k\text{-}1)/q$, $\Delta T(q, 1)=(k\text{-}1)/q+\Delta z$ with $0<\Delta z<k/q\text{-}(k\text{-}1)/q$, and $\Delta T(q, 2)=k/q$.

Fig. 9 shows that multiple partitions appeared in the development of the joint probability resulting from (33) and (34). In the steady state, diagonal matrices are distributed along the diagonal line, where partitions are expressed as the intervals [1, 7], [8, 15], [16, 21], and [22, 30]. In the diagonal matrix area, a high joint probability showing a conspicuous peak is ~1.0 and other joint probabilities are ≈0.0. In the noisy area, all joint probabilities are in the range 0.1–0.3. This pattern of joint probabilities implies a cognitive universe constructed by pasting

diagonal matrix areas in the background of homogeneous noisy areas. This is a “pasted universe.”

Since the symmetric structure between data and hypothesis in the form of (21) and (24) results from Bayes’ formula, someone recognizing a universe through the pasted one could assimilate data with hypotheses (Fig. 10). Fig. 10B shows a matrix expression for the steady-state distribution of the joint probability. If one stays in the diagonal matrix area, one can see a one-to-one correspondence between data and hypotheses. Thus, one can uniquely recognize the corresponding image (*h,* hypothesis) to the given external stimulus (*d,* data) because of the high joint probability with $P(d, h)$. In other words, diagonal matrix area represents a type of attractor. Because such $P(d, h)$ is not strictly equal to 1.0, one cannot remain at the co-ordinate (d, h) indefinitely; one then moves to the second highest peak of the joint probability with $P(d, h')$ (blue arrow in Fig. 10B) or $P(d', h)$ (black arrow in Fig. 10B). The second highest peak exists not in the diagonal matrix area but in the noisy one. That is why the recognition moves from (d, h) to (d, h') or (d', h). Also, due to the assimilation of data and hypotheses, the subsequent transition from (d, h') to (d', h') (blue arrow in Fig. 10B) or from (d', h) to (d', h'') (black arrow in Fig. 10B) can occur. This implies a chaotic transition from one attractor to another, which is argued for in chaotic brain theory (Freeman, 1999, Freeman & Vitiello, 2006, Tsuda, 2002) In that theory, a neuron is regarded as chaotically ryudynamic, in which case a neural network, as a many-degree system, could constitute higher dimensional chaotic dynamics consisting of many attractors. In our model, such a manifold could be generated through cognition.

Neural Net Implementation of BIB inference

In order to manifest the significance of inverse Bayesian inference in neuroscience, we implement Bayesian inference reduced from inverse Bayesian inference in a model of neural networks, i.e., a restricted Boltzmann machine (Smolensky, 1986). An artificial neural network consists of neurons connected with each other by links. The quantitative degree of a connection is expressed as its weight. The time development of an artificial neural network such as the Hopfield model (Hopfield, 1982) is expressed as a transition of the weights and the states of neurons by using the global energy. In contrast to the Hopfield model, a network of Boltzmann machines consists of a visible and a hidden layer as shown in Fig. 11A. In the restricted Boltzmann machine in particular, there is no connection between any two visible units or between any two hidden ones.

The state of the ith visible unit at the tth step is expressed as v_i^t, and that of the ith hidden unit at the tth step is expressed as h_i^t. These states are either 1 or -1. The bias of the ith visible unit at the tth step is expressed as b_i^t, and that of the corresponding hidden unit is

expressed as c_i^t. These biases are real values in the interval [0.0, 1.0]. The weight of the connection between the ith hidden unit and jth visible unit at the tth step is expressed as $w_{i,j}^t$.

The conditional probability of a hidden unit firing under the condition of a visible unit is expressed as

$$P(h_i^t=1|v^t)=\sigma(\sum_{j=1}^{N} w_{i,j}^t v_j^t + c_i^t). \tag{35}$$

The conditional probability of a visible unit firing under the condition of a hidden unit is similarly expressed as

$$P(v_j^t=1|h^t)=\sigma(\sum_{i=1}^{N} w_{i,j}^t h_i^t + b_i^t) , \tag{36}$$

where σ is defined as a sigmoid function such that

$$\sigma(x) = 1/(\exp(-x)). \tag{37}$$

From the derived conditional probability, visible and hidden units firing are collected. Then, the biases and all weights of the connections are updated by the following:

$$c_i^{t+1} = c_i^t + \eta(P(h_i^t=1|v^t)-P(h_i^t=1|v^{t-1})), \tag{38}$$
$$b_j^{t+1} = b_j^t + \eta(v_j^t-vj^{t-1}), \tag{39}$$
$$w_{i,j}^{t+1} = w_{i,j}^t + \eta(P(h_i^t=1|v^t)v_j^t-P(h_i^t=1|v^{t-1})v_j^{t-1}). \tag{40}$$

The variable η represents learning rate and is fixed as 0.005 in all simulating studies here.

These transitions satisfy the algorithm with which we can minimize the Kullback–Leibler divergence. In other words, the minimized state with respect to KL divergence could be achieved in the steady state using this algorithm. Bayesian inference reduced by inverse Bayesian inference is now implanted in the restricted Boltzmann machine by collecting the ith unit such that

$$P(h_k^t=1|v^t)-P(h_k^t=1|v^{t-1})<\theta, \text{ for } k=i \text{ and } i+1. \tag{41}$$

By means of this collection, a set of units is divided into certain partitions. By means of these partitions, Bayesian inference is reduced from relaxation of the unit space. Here, the data

and hypotheses are replaced by hidden and visible units, respectively. Thus, the joint probability of data and hypotheses, $P(d, h)$, is replaced by $w(h, v)$. The reduced Bayesian inference is expressed as

$$w^{t+\Delta T(k,\,1)}(h,\, v_{S(k,\,j)}) = P^{t+\Delta T(k,\,0)}\,(v_{S(k,\,j)}|h), \tag{42}$$

$$w^{t+\Delta T(k,\,2)}(h_{R(k,\,j)},\, v) = P^{t+\Delta T(k,\,1)}(h_{R(k,\,j)}|v), \tag{43}$$

for the visible units of $S(k, j)$ in the kth partition and for the hidden units of $R(k, j)$ in the kth partition, where k=1, 2, …, q, $\Delta T(k, 0)=(k\text{-}1)/q$, $\Delta T(q, 1)=(k\text{-}1)/q+\Delta z$ with $0<\Delta z<k/q\text{-}(k\text{-}1)/q$, and $\Delta T(q, 2)=k/q$, and q is the number of partitions. In (42) and (43), $P^{t+\Delta T(k,\,0)}(h)$ is obtained by summation of $w^{t+\Delta T(k,\,0)}(h, v_{S(k,\,j)})$ for all $S(k, 1)$, $S(k, 2)$, …, $S(k, w_k)$, and $P^{t+\Delta T(k,\,1)}(v)$ is obtained by summation of $w^{t+\Delta T(k,\,1)}(h_{R(k,\,j)}, v)$ for all $R(k, 1)$, $R(k, 2)$, …, $R(k, w_k)$. After the update (38)–(40) for T steps, Bayesian inference reduced by inverse Bayesian inference (42), (43) is applied to the probability distributions of the variable and hidden units.

Figure 11B shows a distribution of the connection weights between visible and hidden units, where a network consists of 200 visible and 200 hidden units, T=3, and the number of updates of the application of (42) and (43) is also three. The weight of the connection between h and v units is represented by a colored dot at the co-ordinate (h, v). The strength of the weight is represented by this color: white, yellow, pink, brown, and black (from weaker to stronger). Influenced by the small difference in the random initial distribution, different partitions develop for each initial condition. Each distribution of connection weight consists of diagonal matrix and homogeneous noisy areas. While the number of applications of Bayesian inference (42), (43) is very small compared to the number of transitions (38)–(40), the perspective of the pasted universe is easily and generally obtained through Bayesian inference reduced from inverse Bayesian inference.

Perspective of the pasted universe

Quantum logic in the form of rough-set driven lattice

In this section, we clarify the significance of the pasted universe consisting of diagonal matrix and noisy areas. We describe the distribution of the joint probability of data and hypothesis in a term of logical structure, and then introduce the lattice driven by a rough set (Pawlak,1991, Järvinen,2007, Yao, 2004, Gunji & Haruna, 2010). Given a universal set U and a map $f:U\rightarrow X$, an equivalence relation R can be defined by

$$(x, y)\in R \quad :\Leftrightarrow \quad f(x) = f(y). \tag{44}$$

The universal set is partitioned into the equivalence classes of R, $[x]_R=\{y\in U|\ (x, y)\in R\}$, which do not overlap with each other. By using this equivalence class, any subset of U can be approximated as a rough set with respect to upper and lower approximations (Pawlak,1991). The upper and lower approximations of X, which is a subset of U, with respect to R are defined by

$$R^*(X) = \{x\in U|\ [x]_R\cap X\neq\emptyset\}, \tag{45}$$

$$R_*(X) = \{x\in U|\ [x]_R\subseteq X\}. \tag{46}$$

Since the equivalence classes are regarded as atoms by which all combinations of atoms can be constructed, collection of subsets of U satisfying $R_*(X) = X$ can constitute a logic in which conjunction and disjunction can be well defined (i.e., Boolean algebra).

If a pair of maps (i.e., two equivalence relations, R and S, for a universal set) is adequately chosen, any lattice can be expressed as a collection of fixed points, L, such that

$$L = \{X\subseteq U|R_*S^*(X) = X\}, \tag{47}$$

where the order in L is defined as an inclusion relation (Gunji & Haruna, 2010). A lattice is an ordered set closed with respect to two binary operations, join and meet (see Appendix), i.e., a type of algebra, and can be compared to logic (Davey & Priestley, 2002). Classical propositional logic, intuitionistic propositional logic, and quantum logic can be compared to Boolean, Heyting, and Orthomodular lattice, respectively. If a fixed point $R_*S^*(X) = X$ is replaced with $S_*R^*(X) = X$, $R^*S_*(X) = X$, or $S^*R_*(X) = X$, the derived lattice is isomorphic to the lattice obtained as (47).

Note that two equivalence relations corresponding to two maps can be interpreted as a one-to-many-type mapping of measurement. Superposition of two maps implies a one-to-many-type mapping. When the distribution of joint probability is analyzed by the rough set lattice, one can estimate the types of one-to-many-type mapping. In the previous section, a distribution of the joint probability of data and hypothesis is expressed as a matrix which consists of diagonal-matrix and noisy areas. If a particular threshold value is introduced to digitize the joint probability, $P(d, h)$; If $P(d, h)$>threshold value, then $(d, h)\in I$. Otherwise, $(d, h)\notin I$, and a distribution of the joint probability is expressed as a binary relation, I, between data and hypothesis. Since this binary relation can be interpreted as the relation between two equivalent classes of equivalence relations, R and S, one can obtain a lattice from the binary relation between data and hypothesis.

As shown in Fig. 12, a lattice corresponding to a given binary relation can be obtained. Given a binary relation such as Fig. 12A, columns and rows can be regarded as two kinds of partitions for a universal set. For a partition {*a, b, …, e*}, each element is an equivalence class derived from a particular map. Similarly, {*A, B*, *C*, *D*} from another map. For *a* in the first partition, *B* and *D* in the second partition have a relation, i.e., (*a, B*)$\in$*I* and (*a, D*)$\in$*I*. Generally, (*g, h*)$\in$*I* implies that there exists an element *p* in the universal set such that *p*$\in$*g* and *p*$\in$*h*. Thus, it implies there are at least two elements in the equivalence class *a*, of which one element belongs to the equivalence class *B* and the other belongs to *D*. Analogously, an equivalence class *e* contains at least three elements of a universal set. Thus, a minimal model for a universal set that can be partitioned both to {*a, b, …, e*} and {*A, B*, *C*, *D*} is obtained as shown in Fig. 12B. Two partitions can be interpreted as a set of equivalence classes that are derived from a pair of maps, *f* and *g*, respectively. Finally, two partitions for a universal set are shown in Fig. 12C. We call the equivalence relation leading to a partition, {*a, b,* …, *e*}, the relation, *S*, and {*A, B, C, D*} the relation, *R*, respectively.

Next, subsets of *U* that satisfy the fixed point $R_*S^*(X) = X$ are collected. It is easily verified that only a union of equivalence classes can be a fixed point and that some unions cannot be fixed points (Yao, 2004, Gunji & Haruna, 2010). All what one has to do to obtain a lattice is to check whether a union of equivalence classes (i.e., a subset of {*A, B, C, D, E*}) can be a fixed point or not. For an empty set, $R_*S^*(\varnothing) = R_*(\varnothing) = \varnothing$ since there is no equivalence class that has an intersection with and contains an empty set. For a singleton set, {*A*}, $R_*S^*(\{A\}) = R_*(\{b, c, e\}) = \{A\}$. Because for *A*, (*b, A*), (*c, A*), (*e, A*)$\in$*I*, $S^*(\{A\}) = \{b, c, e\}$. Because for all elements in {*b, c, e*}, (*b , A*), (*c, A*), (*e, A*)$\in$*I*, $R_*(\{b, c, e\}) = \{A\}$. All singleton sets are analogously fixed points. In contrast, $R_*S^*(\{A, B\}) = R_*(\{a, b, c, d, e\}) = \{A, B, C, D\}$, and then {*A, B*} is not a fixed point. Figure 12D shows a Hasse diagram for an obtained lattice. In a Hasse diagram, each fixed point is represented by a circle. A fixed point *X* included by another fixed point *Y* is connected to *Y* by a line, where the circle representing *X* is located below the one representing *Y*. An obtained lattice is called a rough-set driven lattice (Gunji & Haruna, 2010).

In the previous sections, we showed that the distribution of joint probability, *P*(*d, h*), developed through BIB inference is expressed as a pasted universe consisting of diagonal matrix areas and homogeneous noisy areas. We show that a pasted universe is expressed as an orthomodular lattice. Figure 13A shows a typical binary relation developed by BIB inference, which consists of diagonal relations (i.e., $(d_i, h_i)\in A$ and $(d_i, h_j)\notin A$ with $i \neq j$) and product relations (i.e., for any *i, j*, $(d_i, h_j)\in A$). If $\{d_1, ..., d_5\}$ and $\{h_1, ..., h_5\}$ are considered as sets of equivalence classes, *R* and *S*, respectively, a collection of *X* such as $R_*S^*(X) = X$ can be a lattice.

If a diagonal relation such that $(d_i, h_i) \in A$ and $(d_i, h_j) \notin A$ with $i \neq j$, for i, j = 1, 2, 3, is isolated from a whole relation as shown in Fig. 13A, it is easy to see that the isolated relation can correspond to a Boolean lattice. For a singleton set, $R_*S^*(\{h_1\}) = R_*(\{d_1\}) = \{h_1\}$, and any other singleton sets are also fixed points. For a two-element set, $R_*S^*(\{h_1, h_2\}) = R_*(\{d_1, d_2\}) = \{h_1, h_2\}$, and any other two-elements set are also the sets satisfying the fixed point condition. Thus, it is verified that $L = \{X \subseteq U | R_*S^*(X) = X\}$ is nothing but a power set of $\{h_1, h_2, h_3\}$, which is Boolean algebra.

We now consider the fixed points in a whole relation as shown in Fig. 13A. For a singleton set in $\{h_1, h_2, h_3\}$, $R_*S^*(\{h_1\}) = R_*(\{d_1, d_4, h_5\}) = \{h_1\}$. Although $S^*(\{h_1\}) = \{d_1, d_4, d_5\}$ because of the sub-relation that is the product relation, there is no element h except for h_1 in $\{h_1, ..., h_5\}$ such that for some $d \in \{d_1, d_4, d_5\}$, $(d, h) \in I$ and for any $d' \in \{d_2, d_3\}$, $(d', h) \notin I$. Analogously, any other singleton sets in $\{h_1, h_2, h_3\}$ are fixed points. For a singleton set in $\{h_4, h_5\}$, $R_*S^*(\{h_4\}) = R_*(\{d_1, d_2, d_3, d_4\}) = \{h_4\}$. Although $S^*(\{h_4\}) = \{d_1, d_2, d_3, d_4\}$, there is no h except for h_4 in $\{h_1, ..., h_5\}$ such that for some elements d in $\{d_1, d_2, d_3, d_4\}$, $(d, h) \in I$, and $(d_5, h) \notin I$. Thus, $\{d_1, d_2, d_3\}$ never plays a role in taking h. A singleton set $\{h_5\}$ is also a fixed point for the same reason. For a two-element set in $\{h_1, h_2, h_3\}$, $R_*S^*(\{h_1, h_2\}) = R_*(\{d_1, d_2, d_4, d_5\}) = \{h_1, h_2\}$; that is the fixed point. For a three-element set $\{h_1, h_2, h_3\}$ (the greatest subset) in $\{h_1, h_2, h_3\}$, $R_*S^*(\{h_1, h_2, h_3\}) = R_*(\{d_1, d_2, d_3, d_4, d_5\}) = \{h_1, h_2, h_3, h_4, h_5\}$. Thus, $\{h_1, h_2, h_3\}$ is not a fixed point. Analogously, the greatest subset $\{h_4, h_5\}$ in $\{h_4, h_5\}$ is not a fixed point. A subset consisting of elements both from $\{h_1, h_2, h_3\}$ and from $\{h_4, h_5\}$ is not a fixed point. Since $S^*(\{h_1, h_4\}) = \{d_1, d_2, d_3, d_4, d_5\}$, $R_*(\{d_1, d_2, d_3, d_4, d_5\}) = \{h_1, h_2, h_3, h_4, h_5\} \neq \{h_1, h_4\}$. Finally, any subsets of $\{h_1, h_2, h_3\}$ and those of $\{h_4, h_5\}$ are the fixed points, except for $\{h_1, h_2, h_3\}$ and $\{h_4, h_5\}$, and that any subset originating from $\{h_1, h_2, h_3\}$ and $\{h_4, h_5\}$ is not a fixed point. A collection of fixed points contains all subsets of the Boolean lattice corresponding to each diagonal relation, but not a set consisting of different Boolean lattices corresponding to diagonal relations, as shown in Fig. 13B. Thus, the obtained lattice is expressed as a disjoint union of two Boolean lattices except for the least and the greatest element, which is an orthomodular lattice (Gunji et al., 2016).

This idea is verified generally, which implies that a relation between data and hypotheses digitized from the joint probability developed through Bayesian inference derived from inverse Bayesian inference is destined to be an orthomodular lattice. Since the distribution of the joint probability consists of diagonal areas and homogeneous noisy areas, the relation obtained by digitization is generally expressed as shown in Fig. 14 (left). A set of data $D = \{d_1, d_2, \ldots, d_n\}$ is divided here into subsets of $D_1, D_2, \ldots, D_N$, which reveals a partition of D where $d_1, d_2, \ldots$ are renumbered by partition. Thus, $D = D_1 + D_2 + \ldots + D_N$, which implies that D is a disjoint union of D_k (i.e., the intersection of D_i and D_j with $i \neq j$ is empty), where, for

each subset, $D_k = \{d_{k1}, d_{k2}, \ldots, d_{kv}\}$. Similarly, $H = H_1+H_2+\ldots+H_N$. These partitions show that $I_k \subseteq D_k \times H_k$ is a diagonal relation in which $(d_{ki}, h_{ki}) \in I_k$ and $(d_{ks}, h_{kt}) \notin I_k$ with $s \neq t$, and that $J_{ij} \subseteq D_i \times H_j$ is a product relation in which for any $d_{is} \in D_i$ and $h_{jt} \in H_j$ $(d_{is}, h_{jt}) \in J_{ij}$. A given relation I between D and H is a disjoint union of I_k with k=1, 2, .., N and J_{ij} with $i \neq j$.

Under this situation, one can determine whether a subset of H (i.e., union of h_i and h_j) is a fixed point with respect to R_*S^*; it can be checked with respect to the relation I. For any subset of H represented by $\mathrm{Sub}(H) \subseteq H$, it is verified that

$$S^*(\mathrm{Sub}(H)) = \{d \in D \mid \text{For some } h \in \mathrm{Sub}(H), (d, h) \in I\}. \tag{48}$$

For any $\mathrm{Sub}(D) \subseteq D$, it is also verified that

$$R_*(\mathrm{Sub}(D)) = \{h \in H \mid \text{For any } (d, h) \in I,\ d \in \mathrm{Sub}(D)\}. \tag{49}$$

Therefore, for a proper subset of H_k that is represented by $\mathrm{Pub}(H_k)$ and $(d, h) \in I_k$ with $d \in D_k$ and $h \in H_k$,

$$R_*S^*(\mathrm{Pub}(H_k)) = R_*\Big(\mathrm{Pub}(D_k) + \sum_{j \in \{1, .., N\},\, j \neq k} D_j\Big), \tag{50}$$

where the symbols "+" and "Σ" represents a disjoint union, and $\mathrm{Pub}(D_k)$ represents a proper subset of D_k whose set of indices is equal to H_k (e.g., if $\mathrm{Pub}(H_k) = \{h_{ks}, h_{kt}\}$ then $\mathrm{Pub}(D_k) = \{d_{ks}, d_{kt}\}$). Equation (50) is obtained because, for any d in ΣD_j with $j \in \{1, .., N\}$, $j \neq k$, there exists $h \in H_k$ such that $(d, h) \in J_{jk}$ and then $(d, h) \in I$.

Next, we calculate $R_*(\mathrm{Pub}(D_k)+\Sigma D_j)$, and have to search for $h \in H$ such that for any $(d, h) \in I$, $d \in \mathrm{Pub}(D_k)+\Sigma D_j$. It is clear that $h \in \mathrm{Pub}(H_k)$ satisfies the condition. Because any pair of I_i and I_j are mutually disjoint, for any $d_k \in \mathrm{Pub}(D_k)$ there exists J_{ks} with $s \in \{1, .., N\}$, $s \neq k$ such that $(d_k, h_s) \in I$. However, for those $h_s \in H_s$, there exists $d \in \Sigma D_j$ such that $(d, h_j) \notin I$. Thus, there is no $h \in H$ except for $h \in \mathrm{Pub}(H_k)$ that satisfies that for any $d \in \mathrm{Pub}(D_k)+\Sigma D_j$, $(d, h) \in I$. In other words, ΣD_j with $j \in \{1, .., N\}$, $j \neq k$ does not contribute to taking a set by operating R_*. Then,

$$R_*S^*(\mathrm{Pub}(H_k)) = \mathrm{Pub}(H_k)\,. \tag{51}$$

In contrast, for H_k itself,

$$R_*S^*(H_k) = R_*(D) = H, \tag{52}$$

since $S^*(H_k) = D_k + \Sigma D_j$ ($j \in \{1, .., N\}, j \neq k$) $= \Sigma D_j$ ($j \in \{1, .., N\}$) $= D$. Thus, H_k is not a fixed point. Analogously, for any disjoint union of Pub(H_i) and Pub(H_j),

$$R^*S^*(\mathrm{Pub}(H_i)+\mathrm{Pub}(H_j)) = R^*((\mathrm{Pub}(D_i)+ \underset{k\in\{1,..,N\},\, k\neq i}{\Sigma} D_k) \cup (\mathrm{Pub}(D_j)+ \underset{s\in\{1,..,N\},\, s\neq j}{\Sigma} D_s))$$
$$= R^*(D) = H. \tag{53}$$

Thus, subsets consisting of different H_i and H_j are not fixed points. Since it is clear to see that $R^*S^*(H)=H$, and $R^*S^*(\varnothing)=\varnothing$, it is verified that an obtained lattice as a collection of fixed points is a disjoint union of Boolean lattices where the least and the greatest elements are common, and then an orthomodular lattice.

An orthomodular lattice corresponds to a quantum logic in which the distributive law does not hold. While quantum logic reveals microscopic features ambiguous to the local and non-local, the logic obtained here reveals macroscopic features in cognition.

Previously, there were some attempts to verify the orthomodular lattice not by quantum mechanics but by a macroscopic measurement process (Svozil, 1993, Atmanspacher & Graben, 2015, Gunji et al., 2016). We discuss the significance of our model in the Discussion section below.

Macroscopic reality in an Orthomodular Lattice

Information Generation

Instead of a propositional logic, one can utilize a lattice to understand the logical structure. It is well known that a Boolean lattice corresponds to classical propositional logic. Since a Boolean lattice **B** satisfies (i) the distributive law; for any element *x, y, z* in **B**, $x\wedge(y\vee z) = (x\wedge y)\vee(x\wedge z)$, and (ii) the complement law; for any x in **B**, there exists a complement of x, $x^{\perp}$, such that $x\wedge x^{\perp} = 0$ and $x\vee x^{\perp} = 1$, where 0 and 1 are the least and the greatest elements in **B**, respectively, join ($\vee$) and meet ($\wedge$) can be interpreted as disjunction ($\cup$) and conjunction ($\cap$), respectively (see Appendix). A logical operation, implication ($\rightarrow$), can be expressed by using the complement and conjunction in **B**, such that $x\rightarrow y = x^{\perp}\vee y$. A syllogism such that

$$(x\rightarrow y)\cap(y\rightarrow z) \Rightarrow x\rightarrow z \tag{54}$$

can be expressed in **B**, where $\Rightarrow$ is replaced by the order defined in **B**, as

$$(x^{\perp}\vee y)\wedge(y^{\perp}\vee z)\leq x^{\perp}\vee z. \tag{55}$$

Because the distributive law holds in **B**, by applying it to the left-hand terms, $(x^{\perp}\vee y)\wedge(y^{\perp}\vee z) = (x^{\perp}\wedge y^{\perp})\vee(x^{\perp}\wedge z)\vee(y\wedge z) = (x^{\perp}\wedge(y^{\perp}\vee z))\vee(y\wedge z)$. Since $a\wedge b\leq a\leq a\vee c$ and $c\leq a\vee c$, $(a\wedge b)\vee c\leq a\vee c$. Thus, in replacing *a, b*, and *c* with $x^{\perp}$, $y^{\perp}\vee z$, and $y\wedge z$, respectively, one obtains $(x^{\perp}\wedge(y^{\perp}\vee z))\vee(y\wedge z)\leq x^{\perp}\vee(y\wedge z)$. Similarly in *a, b*, and *c* with $x^{\perp}$, *y*, and *z*, one obtains $x^{\perp}\vee(y\wedge z)\leq x^{\perp}\vee z$. Therefore, the inequality (55) holds.

We now consider a syllogism in an orthomodular lattice **O**. Since **O** satisfies the complement law, implication is defined to be the same as that in a Boolean lattice. However, because **O** does not satisfy the distributive law, (55) does not hold for any elements in **O**. Indeed, there is a case in which the reverse transition of the syllogism can appear. In the orthomodular lattice shown in Fig. 13C, one can find the reverse direction of the syllogism. In $(x^{\perp}\vee y)\wedge(y^{\perp}\vee z)$, substituting *x*, *y*, *z* with $a^{\perp}$, *b*, *a*, respectively, leads to

$$(a\vee b)\wedge(b^{\perp}\vee a) = 1\wedge 1 = 1. \tag{56}$$

This replacement in the right-hand terms in (55) leads to

$$a^{\perp\perp}\vee a = a\vee a = a. \tag{57}$$

The reverse direction of the syllogism implies that

$$x\rightarrow z\Rightarrow(x\rightarrow y)\wedge(y\rightarrow z). \tag{58}$$

As to whether this actually makes sense, one can consider that information *y* could appear by which the implication $x\rightarrow z$ is articulated into two implications. Assume *x* and *z* as statements such as

x: putting a hand in a pocket
z: taking a lucky coin in the pocket. (59)

The left-hand terms of (59) imply that if one puts a hand into the pocket then a lucky coin will be pulled out. Now assume *y* as a statement such as

y: there are lucky and unlucky coins in the pocket. (60)

Consider what happens when statement y appears. The implication $x \rightarrow y$ implies that if one puts a hand into a pocket, one feels that there are coins in it that are either lucky or unlucky. The implication $y \rightarrow z$ implies that if one feels that there are these two kinds of coins, one can choose a lucky one by tactile means. This means that one has the ability to choose lucky coins. The appearance of the statement y could entail "the ability to choose lucky coins."

Another example is illustrated by the statements

x: Drought condition
z: First decent rain. (61)

Here $x \rightarrow z$ implies that if drought conditions persist then someday decent rain will fall. Assume y as a statement such as

y:Mr. X calls for rain. (62)

With the appearance of y, $x \rightarrow y$ implies that Mr. X calls for rain during a drought, and $y \rightarrow z$ implies that it rains after Mr. X calls for it to do so. Thus, the appearance of statement y implies the appearance of an apparent ability to summon rain. Information is, therefore, generated by the inverse direction of the syllogism.

Actual resolution to the frame and symbol grounding problems

When artificial intelligence (AI) was first developed, some critical problems were pointed out, such as the frame problem (Dreyfus & Dreyfus, 1988) and the symbol grounding problem (Harnad,1990). Since AI is implemented in a virtual world, the connection between AI and the real world is lost. It is assumed that logical operations can be well defined in a virtual world and that AI can use such logical operations and symbols adequately. Since AI can use symbols adequately, it can learn the relationship between a symbol in the virtual world and its corresponding object in the real word. Thus, AI can find that "STRIPE" implies the pattern of stripes and "HORSE" implies a particular animal. Since AI can use logical operations adequately, it finds that STRIPE is true (i.e., STRIPE = 1), HORSE is true (i.e., HORSE = 1), and then STRIPE AND HORSE is true (i.e., STRIPE$\wedge$HORSE = 1). Although AI can manipulate symbols and logical operations, it cannot find the emergent grounding between STRIPE AND HORSE and the corresponding "zebra". That is the symbol grounding problem.

The frame problem arises in determining the relationship between a symbol in the virtual world and its corresponding object in the real world. As mentioned in the symbol

grounding problem, at first the relationship between a symbol and its corresponding object is simply assumed, which results in STRIPE being true. Because a stripe is a visual pattern, it is true only if the pattern can be visualized. If it is too dark to be seen, then STRIPE is not true. Therefore, in order to be able to state that STRIPE is true under a general condition, one has to check whether STRIPE is true under any condition, e.g., a cloudy dark day, a rainy day, indoors. The situation is summarized as follows. Replace STRIPE, cloudy day, rainy day, … with symbols A, C_1, C_2, …, respectively. If one at first sees the stripe pattern on the cloudy day (C_1), the situation is expressed as

$$A\wedge(C_1\vee C_2\vee\ldots)\ . \tag{63}$$

We concentrate here on the logical statement as a necessary condition for the real situation, which reveals that the real situation implies the logical statement. With respect to (63), one who sees A under situation C_1 accepts $A\wedge C_1$. That is why (63) holds for the necessary condition for the real situation. In contrast, in order to state that "STRIPE is true" for any conditions, one has to verify

$$A\wedge(C_1\wedge C_2\wedge\ldots)\ . \tag{64}$$

Although empirically (63) is possible but (64) is impossible for any condition, (64) has to be possible logically. That is the essential feature of the frame problem.

By using logical operations, the symbol grounding problem itself can be expressed as the torsion between AND ($\wedge$) and OR ($\vee$). When HORSE is grounded with a real horse or STRIPE is grounded with a real pattern of stripes, it can be stated that

$$\text{HORSE}\ \vee\ \text{STRIPE} \tag{65}$$

is grounded or is true in the real world. Actually, if either HORSE or STRIPE is grounded, then (65) holds. When the HORSE and STRIPE is grounded with a real zebra,

$$\text{HORSE}\ \wedge\ \text{STRIPE} \tag{66}$$

is grounded. Thus, the symbol grounding problem is also expressed as a statement for which the AND expression (66) has to be verified under a condition in which not (66) but just the OR-expression (65) is possible. That is equivalent to the frame problem.

For both the symbol grounding and frame problems, the orthomodular lattice could yield actual solutions. Recall Fig. 13C as an example of an orthomodular lattice. Since an orthomodular lattice never satisfies the distributive law such that $x\wedge(y\vee z) = (x\wedge y)\vee(x\wedge z)$, "$\wedge$" cannot be interpreted as AND, and "$\vee$" cannot be interpreted as OR. However, we here assimilate "$\wedge$" and "$\vee$" with AND and OR, respectively, in an orthomodular lattice. When the assimilation is represented by "~", we obtain

$$A\wedge(C_1\vee C_2) \sim (A\wedge C_1)\vee(A\wedge C_2). \tag{67}$$

In substituting A, C_1, and C_2 by elements a, $b^{\perp}$, and $c^{\perp}$, respectively, the left-hand term becomes $a\wedge(b^{\perp}\vee c^{\perp}) = a\vee 1 = a$. By contrast, the right-hand term is expressed as $(a\wedge b^{\perp})\vee(a\wedge c^{\perp}) = 0\wedge 0 = 0$. Thus, one can obtain $0 = (a\wedge b^{\perp})\wedge(a\wedge c^{\perp}) = a\wedge(b^{\perp}\wedge c^{\perp})$. This implies that in an orthomodular lattice, one can obtain

$$A\wedge(C_1\vee C_2) \sim A\wedge(C_1\wedge C_2). \tag{68}$$

That is an actual solution for the symbol grounding and/or frame problem, due to the orthomodular lattice. Illogical assimilation of conjunction and disjunction can give rise to the solution of the symbol grounding problem and/or frame problem.

The symbol grounding and frame problems could appear in the contact between the virtual and real universes. It is assumed that formal logic and the method of manipulating symbols are invariant and well defined, while objects and phenomena in the real world are dynamic and ambiguous. Thus, the discrepancy between the virtual and real worlds can entail the problem. Since the virtual and formal world is well defined and rigid, the solution could be expected and constructed as the interface between the rigid virtual world and the ambiguous real one. Such solutions can be compared to the solutions in robotics and neuroscience, as mentioned previously (Pfeifer & Scheier, 2001, Brooks,1986; 1991, Pfeifer et al., 2007, Varela et al., 1991, Varela, 1997, Pfeifer & Gomez, 2009). The influences from the real world cannot directly reach the real one, and those from the virtual world cannot directly reach the virtual one. An interface such as the rubber skin for the robotic finger can mediate between the virtual world and the real one (Fig. 15A).

The assimilation of AND (conjunction) with OR (disjunction) such as (68) shows a different solution for the symbol grounding and frame problems. The "well-defined-ness" of formal logic is no longer maintained since the real world can microscopically influence the virtual world anywhere. The discrepancy between the virtual and real worlds cannot be received by the front like an interface as shown in Fig. 15A, which could break the well-

defined conjunction and/or disjunction as shown in Fig, 15B. Thus, conjunction and disjunction can no longer be separated from each other, and must be assimilated instead. That could solve the symbol grounding and frame problems. In other words, the conflict between virtual and real worlds could not only establish the symbol grounding and frame problems but also yield solutions to them (Fig. 15B). Such conflicts could seep through individual logical operations. Well-defined-ness could no longer remain, even in a virtual world.

Discussion

We here implement an aspect of measurement, observation, and cognition based on inference and memory, i.e., BIB inference. This idea is inconsistent with naturalistic dualism for consciousness, in which there is a distinct separation between inside and outside of an observer, and in which it is assumed that an object outside is mapped into a representation inside. Instead, any object or phenomenon outside appears through measurement, and any qualia, feeling or emotion inside also appears through measurement. Anything both inside and outside could exist, potentially accompanied with measurement. That is nothing but objects that appear through internal measurement or endoperspective.

Naturalistic dualism is an issue to be overcome, while the intrinsic difference between mind and matter has to be implemented. Regarding this problem, many philosophers have recently converged to neutral monism, in which a matter-like aspect and a quality-like one could interact dynamically with each other and could give rise to particular configurations (Russel, 1921, Strawson, 2006, Silberstein & Chemero, 2015). When the matter-like aspect is dominant in the configuration, the interaction could be regarded as “matter”, and when the quality-like aspect is dominant, that could be regarded as “mind.” The endoperspective is consistent with neutral monism, or could be an implementation of neutral monism in science. However, the intrinsic difference between mind and matter, or between quantity and quality, was not implemented in neutral monism and endoperspective.

We consider the issue of the qualia, especially that is discussed in the form of philosophical zombie (Chalmers, 1996) is an intrinsic property of consciousness. The philosophical zombie has no qualia while he or she can do as well as human being not only externally (e.g. talking about zombie) but internally (e.g., brain wave). Thus science cannot distinguish human being from philosophical zombie. While this definition is ill-defined, Chalmers declares that the problem regarding consciousness is immediately related to infinite regression of description. Qualia escapes as soon as you feel catching up with what

the qualia is. We call it this kind of property the externality of consciousness. Any other issues of consciousness, for example postdiction and/or intentional consciousness, can be considered without the externality of consciousness. That is why qualia is the special issue not only in philosophy but in science of consciousness.

A pair of Bayesian and inverse Bayesian inference implements the forward (from premise to conclusion) and backward (from conclusion to premise) reasoning. This pair is the main engine to bring the outside implicitly. The axiomatic metric space is based on the notion of metric, and it verifies the notion of open sets. In this sense metric is premise, and the notion of open sets is conclusion. That is forward reasoning. In history of mathematics, backward reasoning is introduced. The notion of open set is regarded as axioms and topological space is defined. Then, it is verified that metric space is a part of topological space. In this sense open set is premise, and metric space is conclusion. Due to the transition from the forward reasoning to the backward reasoning the outside of the metric space is implicitly imported in topological space. As well as forward and backward reasoning, Bayesian and inverse Bayesian inference can contribute implicit importing the outside. That is why these inferences are immediately relevant for the externality of consciousness, qualia.

Measurement and observation could be expressed as a map from an objective universe outside to a subjective one inside. As soon as such a separation between the inside and outside is accepted, it is cancelled. That could be an implementation of the endoperspective. Such a cancellation is destined to contain a discrepancy between the inside and outside. In our proposed inference system, the relationship between the inside and outside is replaced with that between parts and a whole and between the smaller and the larger parts of the probability space. Since it is assumed that the whole of the probability space is assimilated with the real universe, it is regarded as the universe outside. Once a whole space is replaced with a subspace dependent on a particular context, or the larger space is replaced with a smaller one, the objective universe outside is replaced with the subjective universe dependent on a particular condition. However, such a replacement is only possible as far as the replacement can be iterated. The iteration of replacing a part with the whole also reveals replacing the whole with a part, and then gives rise to interpolation of parts and whole. In this sense, the relationship between parts and whole can be embedded with the intrinsic discrepancy existing between the inside and outside.

Bayesian inference is an implementation in which the outside is perpetually replaced with the inside. A probability that is independent of any condition is replaced with a conditional probability under a particular condition. Special cases are continually generalized. Given a probability space, a system focuses only on the subspace related to a given particular condition and ignores other subspaces by changing the distribution of the probability. This

can result in immediately approaching the optimal solution for a given particular condition. It implies that the system can immediately reach the optimal solution for an external stimulus. That is why such an inference system (e.g., a human) can immediately make a decision fit to the condition.

There is no reverse directed inference, of which the inside is replaced with the outside, only in Bayesian inference system. Our inverse Bayesian inference is such an implementation, in which the conditional probability dependent on the condition (hypothesis) is continually replaced by the probability of the empirical data derived from the universe outside. Actually, the conditional probability dependent on a hypothesis that is ignored by Bayesian inference is replaced by the probability of the empirical data. This implies that hypotheses stored in the system are replaced continually by new ones. In this sense, the inside is constantly replaced with the outside. If the external condition is stable and the probability distribution of the data is unchanged, the inverse Bayesian inference cannot contribute to the system's decision making. Since newly introduced hypotheses are replaced with hypotheses with low probability, those new hypotheses are also labeled with low probability. Thus, they cannot be used in making decisions. In contrast, if the external condition is unstable and the probability distribution of data is continually changed, then the distribution of the probability of hypotheses is also continually changed. This implies that even hypotheses with the least probability can be changed those with the greatest hypothesis that is used to make a decision. The inverse Bayesian inference, therefore, plays a role in making a decision.

The probability space of hypotheses is contracted by Bayesian inference and is relaxed (expanded) by inverse Bayesian inference. The measurement process with respect to inference has not been implemented until the two-sided inference is implemented. Conversion of parts and the whole is a contradiction if wholeness is invariant. In probability theory, distribution of the probability is invariant through time. In our inference system, not only distribution of the probability but events designating the probability themselves are changed and modified. The former is caused by Bayesian inference and the latter by inverse Bayesian inference. Thus, the notion of changing probability could constitute a discrepancy that is an expression for the intrinsic difference between the inside and outside.

The idea of Bayesian inference is extremely prevalent in cognitive science because it can facilitate immediate decision-making adapted to the external environment (Gigerenzer & Hoffrage, 1995, Knill & Pouget, 2004). In experiments in cognitive science, a particular condition is set for subjects, and their behavior is observed and evaluated. In order to detect well-defined experimental results, the condition has to be stable and to be uniquely determined. That is why cognitive scientists pay attention only to Bayesian and not to inverse

Bayesian inference. Neuroscience, however, records and collects any neural activity to find out the neural loci correlated with consciousness, and then reaches the hypothesis of global workspace that contains the idea of BIB inference (Dehaene et al., 1998, Dehaene & Naccache,2001, Dehanene & Changeux,2011).

As mentioned before, neural internal selection, by which the largest neural population of synchronized firing is selected, can correspond to Bayesian inference. That is just a strategy that approaches the optimal solution. In the hypothesis of global workspace, selected neural activity is globally propagated to other neural areas and can be used. This passive attitude of being used as an employer sees fit can be interpreted as intentional consciousness. It implies that the probability space is expanded and relaxed, which can correspond to inverse Bayesian inference.

In particular, we define a special expression for BIB inference in the idealized implementation. Bayesian inference, in which joint probability is replaced by conditional probability, is derived from inverse Bayesian inference. Since Bayesian inference is temporally applied to the joint probability distribution of data and hypotheses, it results in the temporal alternation of BIB inference in the inference process.

The most important aspect of this type of BIB inference is universal convergence of the pasted universe (i.e., an orthomodular lattice corresponding to quantum logic) consisting of multiple Boolean algebras. For any randomly distributed joint probability of data and hypotheses, a steady state of the distribution appears immediately in the form of diagonal matrix areas and homogeneous noisy ones. A diagonal matrix area shows a one-to-one relationship between data and hypothesis. This implies that the optimal solution (hypothesis) for input (data) is uniquely determined while it is dependent on the condition. The noisy areas can contribute to the random wandering from one diagonal matrix area to another because the peak in the diagonal matrix area is not labeled with unit probability. Instead, it can move to a noisy area with a certain probability. This behavior can reveal chaotic itinerancy in brain theory (Freeman, 1999, Freeman & Vitiello, 2006, Tsuda, 2002). There are certain attractors in the manifolds of neural dynamical systems, and the behavior of neurons is affected by a particular attractor. It remains there for a while, and then suddenly and randomly escapes from that attractor to move to another one. Attractors and trajectories between attractors in chaotic itinerancy can be compared to the noisy and diagonal matrix areas.

Without the assumption of complementarity, our inference model can give rise to a non-distributive or orthomodular lattice. Measurement models in endophysics were previously proposed by introducing incomplete knowledge for objects (Atmanspacher, 2003, Svozil,1993, Atmanspacher & Graben, 2015). Automata are defined by sets of input and output symbols, a set of internal states, the output function, and the transition rule for the

internal state. If an observer gets no information about the internal state of a particular automaton but he can observe and collect output of automata for given inputs for each internal state, then he obtains multiple output functions for each internal state (Svozil,1993). Since the equivalence classes, which are atoms to explain the behavior of the automaton, can be derived from a map, a set of input symbols can be partitioned by each map. When the multiple partitions are pasted in identifying an equivalence class with elements of input symbols, a structure of a partially ordered set is obtained to reveal logic employed to an observer with incomplete information. However, only a specially defined automaton can show a lattice (algebraic structure or logic), and only a more specialized automaton in which complementarity is implemented in partitions can show an orthomodular lattice (Svozil,1993). Thus, an orthomodular lattice has no universality.

If anything appears via measurement, it has to be described in measurement-oriented logic such as an orthomodular lattice, even in a macroscopic perspective, cognition, and/or perception. That is a goal of endophysics. However, a non-distributive orthomodular lattice can appear only if the principle of complementarity is implemented for two measurement systems. Recently, a toy model for measurement proposed in the form of a firefly-box thought experiment (Foulis,1999) was described in terms of lattice theory (Atmanspacher & Graben, 2015). Since complementarity in measurement is assumed, an obtained lattice is constructed by pasting together two Boolean lattices to form an orthomodular one. Complementarity is not the result of phenomena but an intrinsic principle in those models.

By contrast, in our inference system consisting of BIB inference, orthomodularity is universally generated without the principle of complementarity. Only an alternation of contraction and relaxation of the probability space could lead to the orthomodular lattice. As a result, a logical space in which multiple Boolean lattices are pasted in a whole lattice can be generated. There is no mechanism similar to complementarity to make multiple Boolean algebras without a degree of overlapping. The complementarity here is not the *a priori* intrinsic principle but the *a posteriori* results due to the measurement process based on inference. That is the essential property of macroscopic phenomena viewed from the endoperspective. Quantum logic results not from quantum mechanics but from macroscopic properties in measurement.

If there is no inverse Bayesian inference, only a single Boolean algebra is obtained. That is a one-to-one relation between input and output, and a set of atoms by which phenomena can be reduced (Shinohara et al., 2007). In terms of joint probability, there is one hypothesis for a datum that satisfies its joint probability, and it is ~1.0. A whole distribution of the joint probability is expressed as a diagonal matrix. In an orthomodular lattice, however, a Boolean lattice holds under a restricted condition. If multiple elements are taken from one

Boolean sub-lattice, one can verify the distributive law for expressions of lattice polynomial containing the elements and lattice operations. This implies reductionism. However, if some elements are taken from different Boolean sub-lattices, the distributive law no longer holds. Phenomena described in the lattice cannot be reduced to atoms. This non-distributivity entails the reverse direction of syllogism and information generation, and can yield non-logical solutions to the symbol grounding and frame problems.

When we showed previously that BIB inference could give rise to an orthomodular lattice (Gunji et al., 2016), Bayesian inference (contraction) and inverse Bayesian inference (relaxation) were compared to the lower and upper approximations in rough set theory. Compared to a given set, the lower approximation is smaller and the upper one is larger. Superposition of two kinds of approximation can reveal an orthomodular lattice. In this article, we directly implement BIB inference by using probability. The inverse Bayesian inference can play a role in making an orthomodular lattice.

Since Chalmers declared understanding consciousness and/or qualia to be a hard problem, various researchers have adopted phenomenal consciousness in fields such as neuroscience, cognitive science, and robotics. Although the idea of phenomenal consciousness has the potential to embed subjective quality in matter, such a quality was interpreted as a property that cannot be separated from the body, its surroundings, and the whole universe. The singularity and locality of the subjective quality was then lost in the notion of embodiment. Philosophers who focused on the singularity of qualia and consciousness moved toward panpsychism. When this failed, many philosophers, including analytic ones from the philosophy of science, moved toward panqualityism and/or neutral monism. Viewed from the perspective of science, however, this appeared to be a form of spiritualism.

The only way to connect science to neutral monism in separating spiritualism is the endoperspective. Nothing that is independent of measurement is a real entity. Anything is destined to be accompanied with a hidden measurement process. If so, tokens of observation and/or an observer with subjective quality could be embedded in matter. One problem remains: non-distributivity could not be explained without a particular restriction that can be expected in a macroscopic universe in which partial reductionism and analytical methods are useful and non-reductionism-complexity is observed as a whole. Our measurement system, consisting of BIB inference, solved this problem by leading to the universality of orthomodularity.

Finally, we refer to the qualia lattice mentioned in (Balduzzi & Tononi, 2009). Tononi mentioned a stable subnetwork as qualia and called a set consisting of all combinations of qualia (atoms) a qualia lattice. Due to this definition, his qualia lattice is destined to be Boolean. In our model for a given time series of an external stimulus, BIB inference

contributes to generate an orthomodular lattice, where subjective feeling could wander over the atomic components (i.e., diagonal matrix areas). This kind of dynamic wandering can be attributed to subjective qualia. An orthomodular lattice can be regarded as a qualia lattice.

Acknowledgement

Our research was financially supported by JSPS 15K12054.

Appendix

An ordered set and a lattice are defined by the following.

Definition 1 (Order relation)

Given a set S, let x, y, and z be elements of S. A binary relation $R \subseteq S \times S$ satisfying the following condition is called an order relation.

(i)	xRx	(reflective law)
(ii)	$xRy,\ yRx \Rightarrow x = y$	(anti-symmetric law)
(iii)	$xRy,\ yRz \Rightarrow xRz$	(transitive law)

Order relation xRy can be expressed as $(x, y) \in R$ and also $x \leqq y$. $x \leqq y$ can be expressed as $y \geqq x$. A set equipped with order relation is called an ordered set.

Definition 2 (Meet and Join)

Given an ordered set S, take M that is a subset of S. If $m \leqq a$ for every element m in M, a is called an upper bound for M. A set of upper bounds for M is represented by M^u. Similarly, $b \leqq m$ for every element m in M, b is called a lower bound for M. A set of lower bounds for M is represented by M^l.

The least upper bound for M is called a join for M and is represented by $\vee M$, if it exists, and the greatest lower bound is called a meet for M and is represented by $\wedge M$, if it exsits.

From the definition of join and meet, it is clear that

$$\forall m \in M,\ m \leqq a \Rightarrow \vee M \leqq a$$

$$\forall m \in M,\ b \leqq m \ \Rightarrow\ b \leqq \bigwedge M.$$

When M is a two-element set such that $M = \{x, y\}$, A join $\bigvee M$ is expressed as $x \vee y$, and a meet $\bigwedge M$ is expressed as $x \wedge y$

Definition 3 (Lattice)

An ordered set L is a lattice if and only if for every pair of elements $x, y \in L$, a join $x \vee y$, and a meet $x \wedge y$ are also elements of L.

Definition 4 (Distributive Lattice)

A lattice L is a distributive lattice if and only if for all elements $x, y, z \in L$,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

Symmetrical equation $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ can be verified straightforwardly from $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

Definition 5 (Complemented Lattice)

A lattice L is a complemented lattice if and only if for any elements $x \in L$, there exists the complement of x, $x^{\perp}$, such that

$$x \wedge x^{\perp} = 0 \text{ and } x \vee x^{\perp} = 1,$$

where 0 and 1 are the least and the greatest element in L.

Definition 6 (Boolean Lattice/ Boolean algebra)

A lattice that is both distributive and complemented is called a Boolean lattice or a Boolean algebra.

Definition 7 (Orthomodular Lattice)

A lattice L is an orthomodular lattice if and only if for all elements $x, y, z \in L$,

$$x \geqq y \;\Rightarrow\; y = x \wedge (y \vee x^{\perp})$$

where $x^{\perp}$ is a complement for x.

References

Arecchi FT (2003) Chaotic neuron dynamics, synchronization and feature binding: quantum aspects. Mind and Matter: 1, 15-43.

Arecchi FT (2011) Phenomenology of Consciousness: from Apprehension to Judgment. Nonlinear Dynamics, Psychology and Life Sciences 15**:** 359-375.

Atmanspacher H, Römer H, Walach H (2002) Weak quantum theory: complementarity and entanglement in physics and beyond. Found Phys 32: 379-406.

Atmanspacher H (2003) Mind and matter as asymptotically disjoint, inequivalent representations with broken time-reversal symmetry. BioSystems 68: 19-30.

Atmanspacher H, Graben PB (2015) Complementary observables and non-Boolean logic outside quantum physics. arXiv preprint arXiv:1510.03325, 2015 - arxiv.org.

Balduzzi D, Tononi G (2009) Qualia: the geometry of integrated information. PLoS Comput Biol 5: e1000462.

Brooks RA (1986) A robust layered control system for a mobile robot. IEEE J Robotics and Automation RA-2: 14-23.

Brooks RA (1991) Intelligence without representation. Artificial Intelligence 47: 139-159.

Chalmers DJ (1996) Conscious Mind: In Search of a Fundamental Theory. New York NY: Oxford University Press.

Chalmers D (2007) Naturalistic dualism. In: Velmans M, Schneider S, editors. The Blackwell Companion to Consciousness. London: Blackwell Pub Ltd, 359-368.

Clark A (1998) Being There: Putting Brain, Body, and World Together Again. Cambridge MA: The MIT Press.

Chalmers DJ (2015) Panpsychism and panprotopsychism. In: Alter T, Nagasawa Y editors. Consciousness in the Physical World: Perspectives on Russelian Monism. Oxford Univ Press, 246-276.

Clark A (2003) Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence. New York NY: Oxford University Press.

Couzin ID, Krause J, James R, Ruxton GD, Franks, MR (2002) Collective memory and spatial sorting in animal groups. J Theor Biol 218: 1–11.

Coleman S (2012) Mental chemistry: Combination for panpsychists. Dialectica 66: 137-66.

Davey BA, Priestley HA (2002) Introduction to Lattices and Order. Cambridge: Cambridge University Press, second edition.

De Jaegher H, Di Paolo E, Gallagher S (2010) Can social interaction constitute social cognition? Trends in Cognitive Sciences 14: 441–447.

Dehaene S, Kerszberg M, Changeux JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci USA 95: 14529–14534.

Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition 79: 1-37.

Dehaene S, Changeux J-P (2011) Experimental and theoretical approaches to conscious processing. Neuron 70: 200-227.

Dreyfus H, Dreyfus S (1988) Mind Over Maachine, New York: Free Press.

Foulis DJ (1999) A half-century of quantum logic. What have we learned? In Aerts D edior. Quantum Structures and the Nature of Reality. Dordrecht: Kluwer Press. 1–36.

Freeman WJ (1999) How brains make up their minds. London: Weidenfeld & Nicolson.

Freeman WJ, Vitiello G (2006) Nonlinear brain dynamics as macroscopic manifestation of underlying many-body field dynamics. Phys Life Rev 3: 93–118.

Frith CD, Blakemore S-J, Wolpert DM (2000) Abnormalities in the awareness and control action. Phil Trans R Soc Lond B 355: 1771-1788.

Gallagher S (2000) Philosophical conceptions of the self: implications for cognitive science. Trends Cogt Sci 4: 14–21.

Gallagher S, Zahavi D (2008) The Phenomenological Mind, London: Routledge.

Gigerenzer G, Hoffrage U (1995) How to improve Bayesian reasoning without instruction: Frequency formats. Psychol Rev 102: 684-704.

Grush R, Churchland PS (1995) Gaps in Penrose’s toiling. J Cons Stud 2: 10-29.

Gunji YP (1994) Autonomic life as the proof of incompleteness and Lawvere's theorem of fixed point. Appl Math Comp 61: 231-267.

Gunji YP, Toyoda S (1997) Dynamically changing interface as a model of measurement in complex systems. PhysicaD101: 27-54.

Gunji YP, Ito K, Kusunoki Y (1997) Formal model of internal measurement: alternate changing between recursive definition and domain equation. PhysicaD 110: 289-312.

Gunji YP, Kamiura M (2004) Observational heterarchy enhancing active coupling. Physica D 198: 74-105.

Gunji YP (2004) Proto-Computation or Ontological Measurement. Tokyo: Tokyo University Press (in Japanese).

Gunji YP, Haruna T (2010) A Non-Boolean Lattice Derived by Double Indiscernibility. Transactions on Rough Sets XII: 211-225.

Gunji YP, Sonoda K, Basios V (2016) Quantum cognition based on an ambiguous representation derived from a rough set approximation. BioSystems 141: 55-66.

Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. Nature Neuroscience 5: 382-385.

Hameroff SR, Penrose R (1996a) Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. Mathematics and Computers in Simulation 40: 453-480.

Hameroff SR, Penrose R (1996b) Conscious events as orchestrated space-time selections. J Cons Stud 3: 36-53.

Harnad S (1990) The symbol grounding problem. Physica D 42: 335-346.

Heidegger M (1927=1996) Being and Time. Macquarrie J, Robinson E (trans) New York NY: Harper and Row (='Sein and Zeit').

Hopfield JJ(1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79: 2554-2558.

Husserl E (1913=2001) Experience Unbound. 317 Logical Investigations, 3 vols. Findlay JN (trans.) London, Routeledge.

Igamberdiev AU, Shlovskiy-Kordi NE (2016) Computational power and generative capacity of genetic systems. BioSystems 142; 1-8.

Järvinen J (2007) Lattice theory for rough sets. Transactions on Rough Sets IV, Lecture Notes in Computer Science 4374, 400-498.

Jibu M, Hagan S, Hameroff SR, Pribram KH, Yasue Y (1994) BioSystems 32: 195-209.

Johansson P, Hall L, Sikström S, Olsson A (2005). Failure to detect mismatches between intention and outcome in a simple decision task. Science 310: 116–119.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences 27: 712-719.

Koch C (2012) Consciousness: Confession of a Romantic Reductionist. Cambridge MA: The MIT Press.

Libet B, Greason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. Brain106: 623-642.

Maeno K (2005) How to make a conscious robot: Fundamental idea based on passive consciousness model. Jp Robot Soc 23: 51-62 (in Japanese).

Manktelow K (2012) Thinking and Reasoning: An Introduction to the Psychology of Reason, Judgment and Decision Making. London: Psychology Press.

Matsuno K (1989) Protobiology: Physical Basis for Biology. Boca Raton MI: CRC Press.

Matsuno K, Paton RC (2000) Is there a biology of quantum information? Biosystems 55: 39–

46.
Oizumi M, Albatakis L, Tononi G (2014) From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. Plos Comp Biol 10: e1003588.
Olfati-Saber R (2006) Flocking for multi-agent dynamic systems : algorithms and theory. IEEE Trans Auto Cont 51 : 401–420.
Pawlak Z (1991) Rough Sets, Theoretical Aspects of Reasoning about Data. Boston: Kluwer Academic Pub.
Pfeifer R, Scheier C (2001) Understanding Intelligence. Cambridge UK: MIT Press.
Pfeifer R, Lungarella M, Iida F (2007) Self-organization, embodiment, and biologically inspired robotics. Science 318, 1088-1093.
Pfeifer R, Gomez G (2009) Morphological computation: connecting body, brain, and environment. In: Sendhoff B, Körner E, Sporns O, Ritter H, Doya K, editors. Creating Brain-like Intelligence. Berlin: Springer Verlag.
Popper KR, Eccles JC (1977) The Self and its Brain: An argument for Interactionism. Berlin: Springer Verlag.
Rees G, Kreiman G, Koch C (2002) Neural correlate of consciousness in human. Nature Reviews Neuroscience 3: 261-270.
Reynolds CW (1987) Flocks, Herds, and Schools: A Distributed Behavioral Model. Computer Graphics 21 : 25–34.
Rössler OE (1996) Ultraperspective and endophysics. BioSystems 38: 211-219.
Rössler OE (1998) Endophysics. Singapore World Scientific.
Russel B (1921) The Analysis of Mind. London: George Allen and Unwin.
Scott D (1972) Continuous lattices. In: Lawvere FW, editor. Toposes, Algebraic Geometry and Logic. Berlin: Springer, 97-136.
Seager W (2012) Classical levels, Russelian monism and the implicate order. Found Phys 43: 548-567.
Shinohara S, Taguchi R, Hashimoto T, Katsurada K, Nitta T (2007) Emergence of learning biases and structuring caused by infant agent with symmetry bias. J IPS Japan 48: 125-146.
Silberstein M, Chemero A (2015) Extending neutral monism to the hard problem. J Cons Stud 22: 181-194.
Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. Annual Reviews of Neuroscience 18: 555–586.
Skrbina D (2009) Why neutral monism in superior to panpsychism. Mind & Matter 7: 239-248.
Smolensky P (1986). Information Processing in Dynamical Systems: Foundations of Harmony Theory. In Rumelhart DE, McLelland JL editors. Parallel Distributed Processing:

Explorations in the Microstructure of Cognition, Volume 1: Foundations. The MIT Press. 194–281.

Strawson G (2006) Realistic monism: why physicalism entails panpsychism. J Cons Stud 13; 3-31.

Svozil K (1993) Randomness and Undecidability in Physics. Singapore, World Scientific.

Synofzik M, Vosgerau G, Newen G (2008) Beyond the comparator model: a multifactorial two-step account of agency. Conscious Cogn. 17: 219–239.

Tononi G (2008) Consciousness as integrated information: a provisional manifesto. Biol Bull 215: 216–242.

Tononi G, Koch C (2016) Consciousness: here, there and everywhere? Phil Trans R Soc B 370: 20140167.

Tsakiris M, Prabhu G. Haggard P (2006) Having a body versus moving your body: how agency structures body-ownership. Conscious. Cogn. 15: 423–432.

Tsuda I (2002) Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. Behavioral and Brain Sciences 24: 793-810.

Tye M (1997) Ten Problems of Consciousness: A representational Theory of the Phenomenal Mind. Cambridge MA: The MIT Press.

Yao YY (2004) A comparative study of formal concept analysis and rough set theory in data analysis. Lecture Notes in Computer Science 3066: 59-68.

Varela F, Thompson E, Rosch E (1991) The Embodied Mind, Cambridge, MA: MIT Press.

Varela F (1997) Patterns of life: Intertwining identity and cognition. Brain Cognition 34: 72–87.

Figure Captions

Figure 1. Schematic diagram of an image (object) accompanied with memory. It implies also an object in which memory is embedded. Memory is defined as an interpolating system equipped with contraction and relaxation. Each cause–effect loop expressed as a triangle corresponds to an individual past. Plural pasts are superimposed, which is expressed as memory in the form of a cone. In the framework of Bayesian probability, contraction and relaxation can be replaced by Bayesian inference and inverse Bayesian inference, respectively.

Figure 2. Schematic diagram for Bayesian inference based on inverse Bayesian inference. In this scheme, some hypotheses (data, respectively) are collected and are regarded as all hypotheses, which implies that a part of hypotheses space (data space) is expanded and is relaxed to a whole space. It results in the Bayesian inference of which a joint probability $P(d, h_{s(j)})$ is replaced with conditional probability, $P(d, h_{s(j)})/P_d$ (left diagram), and probability $P(d_{r(j)}, h)$ is replaced with $P(d_{r(j)}, h)/P_h$ (right diagram).

Figure 3. (A) Conditional probability of datum “1” under the optimal hypothesis against time, obtained only by Bayesian inference (green), and a given probability of datum “1” (red). (B) Conditional probability of datum “1” under the optimal hypothesis against time, obtained by both BIB inference (green), and a given probability of datum “1” (red).

Figure 4. (A) Conditional probability of datum “1” under the optimal hypothesis against time, obtained only by Bayesian inference (green), a given probability of datum “1” (red), and accumulated probability of datum “1” (blue). (B) Conditional probability of datum “1” under the optimal hypothesis against time, obtained by both BIB inference (green), a given probability of datum “1” (red), and accumulated probability of datum “1” (blue).

Figure 5. (A) Conditional probability of datum “1” under the optimal hypothesis against time, obtained only by both Bayesian and inverse Bayesian inference, where the hypothesis replaced by empirical data in inverse Bayesian is chosen randomly instead of least optimal hypothesis (red). (B) Conditional probability of datum “1” under the optimal hypothesis against time, obtained by both BIB inference, where the hypothesis replaced by empirical data in inverse Bayesian is the most optimal hypothesis (red).

Figure 6. (A) Probability of hypothesis h_0, h_1, …, h_5 (brown), h_6(blue green), h_7(black), …, h_9 against time, obtained by both Bayesian and inverse Bayesian inference, where the least

optimal hypothesis is replaced by empirical data in the inverse Bayesian inference. (B) Conditional probability of datum “1” under the hypothesis h_6 (blue) and h_7 (red) against time. Other conditions are the same as in A. (C) Conditional probability of datum “1” under each hypothesis (h_0, h_1, …, h_9). Other conditions are the same as in A. (D) Conditional probability of datum “1” under the optimal (red). Other conditions are the same as in A. A given probability of datum “1” (black).

Figure 7. (A) Probability of hypothesis h_0, h_1, …, h_4 (purple), …, h_6 (blue green), h_7 (black), h_8 (gray), h_9 against time, obtained by both BIB inference, where the hypothesis replaced by empirical data is randomly chosen in the inverse Bayesian inference. (B) Conditional probability of datum “1” under the hypothesis h_6 (blue) and h_7 (red) against time. Other conditions are the same as in A. (C) Conditional probability of datum “1” under each hypothesis (h_0, h_1, …, h_9). Other conditions are the same as in A. (D) Conditional probability of datum “1” under the optimal (red). Other conditions are the same as in A. A given probability of datum “1” (black).

Figure 8. Distribution of the joint probability $P(d, h)$ plotted against data, d and hypotheses, h for initial condition (t=0) and for a steady state (t=10).

Figure 9. Distribution of the joint probability $P(d, h)$ plotted against 30 data, d and 30 hypotheses, h for initial condition (t=0) and for a steady state (t=100).

Figure 10. (A) Distribution of the joint probability $P(d, h)$ plotted against data, d and hypotheses, h for a steady state (t=100). (B) Matrix expression for the joint probability where a matrix consisting of red and white cells is a diagonal matrix area and a matrix consisting of orange cells is a noisy area. Arrows represent probabilistic transition from an attractor to another attractor. (C) Diagonal matrix is obtained by the exchange of rows or columns.

Figure 11. (A) Neural networks of restricted Boltzmann machine. The symbol v_i represents a neuron at the visible layer, and h_i represents a neuron at the hidden layer. The symbol $w_{i,j}$ represents the weight of the connection between the ith and the jth neurons. (B) Four examples of a distribution of the joint probability $P(d, h)$ plotted against data, d and hypotheses, h, at t=3. The order from the low to the high probability is represented by the order of the white, yellow, pink, brown, and black dots.

Figure 12. (A) Example for a binary relation between data and hypothesis. A relation is not necessarily symmetric with respect to data and hypothesis. (B) Row and column constituting a binary relation (A) can be regarded as a pair of partitions of a universal set. Thus, a pair of maps, *f* and *g* can be obtained to lead to a pair of partitions. (C) When elements of a universal set is represented by dots, two partitions connecting by a binary relation (A) are expressed as a set of loops (left and right). (D) Lattice obtained from a binary relation (A) in the form of a collection of fixed points of (45).

Figure 13. (A) Typical example for a binary relation between data and hypothesis, obtained from Bayesian inference reduced from inverse Bayesian inference. A relation consists of two diagonal matrix areas and homogeneous noisy areas. (B) Lattice obtained from a binary relation (A) in the form of a collection of fixed points of (45). Because of noisy areas, Boolean sub-algebras are contained in an obtained lattice. (C) An element of the lattice is replaced by a particular element by which orthomodular lattice is evoked.

Figure 14. Binary relation between data and hypothesis derived from joint probability of data and hypothesis developed through BIB inference (left), and its corresponding orthomodular lattice shown as a Hasse diagram (right below). The binary relation consists of sub-relations, diagonal relation, I_k (only diagonal pairs are contained in the relation; left in the right above) and product relation $J_{i,j}$ (all pairs are contained in the relation; right in the right above). The symbol "+" represents the operation of the disjoint union. Each Boolean sub-lattice in a lattice corresponds to the diagonal relation (e.g., $\mathbf{B}_k$ corresponds to I_k). Black circles in a Hasse diagram represent elements of a lattice. White circles connecting the greatest element reveal that the greatest element of each Boolean sub-lattice has the common unique element. White circles connecting the least element also show the similar situation with respect to the least element of the lattice.

Figure 15. (A) Schematic diagram of previously expected solution to the symbol grounding and the frame problem. In sustaining well-defined-ness of formal language and symbol manipulation in virtual world, the interface between virtual and real world is constructed by which the grounding can be mediated. (B) Schematic diagram of our solution in dynamic non-distributive logic to the symbol grounding and frame problem. The influence from the real world is locally and perpetually appeared in the virtual world which could give rise to non-logical mixture of conjunction (AND) and disjunction (OR).

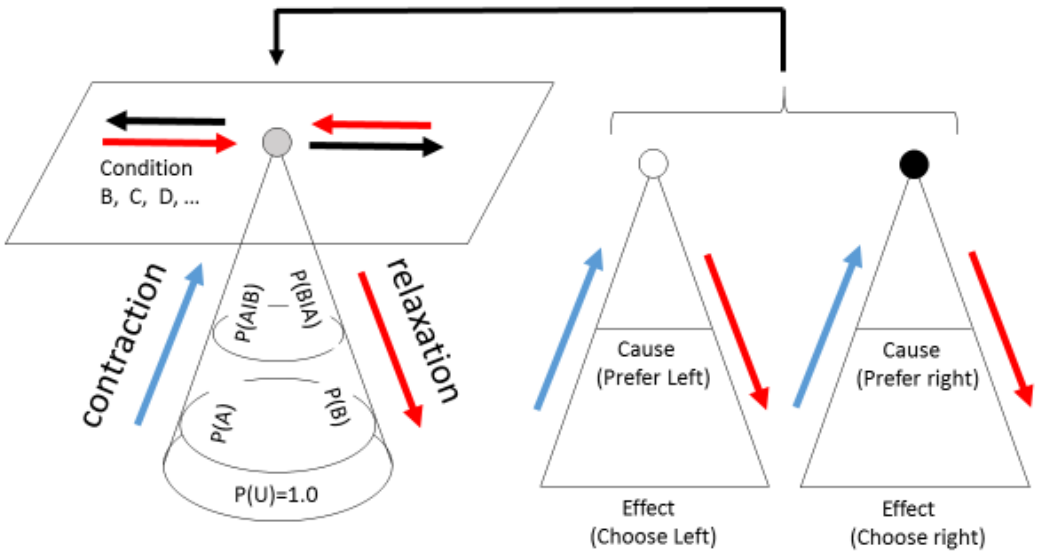

Figure 1

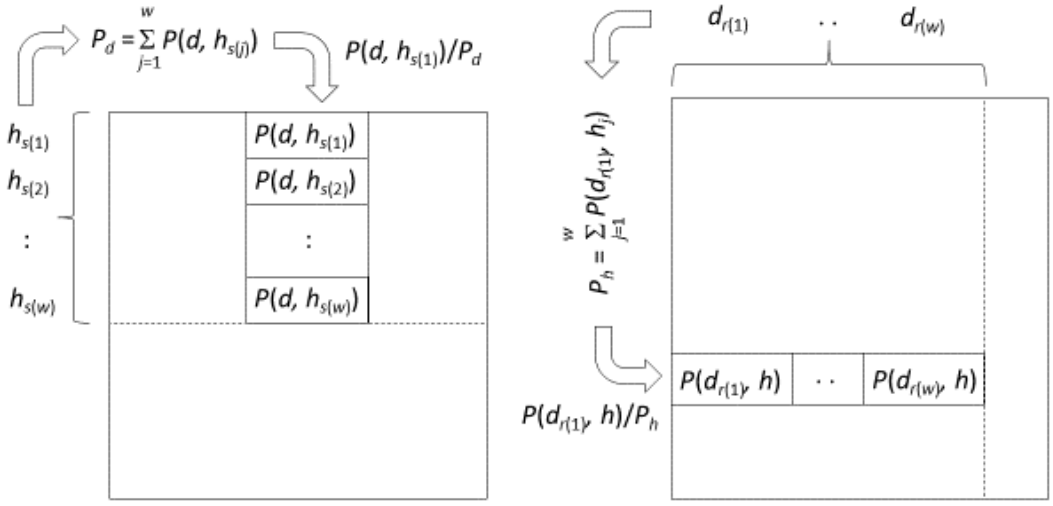

Figure 2

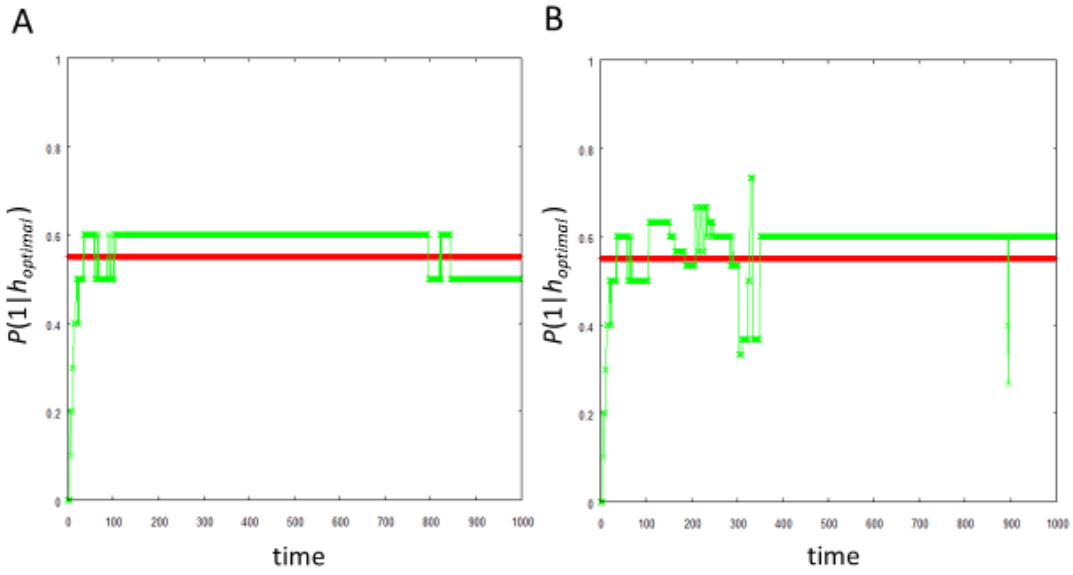

Figure 3

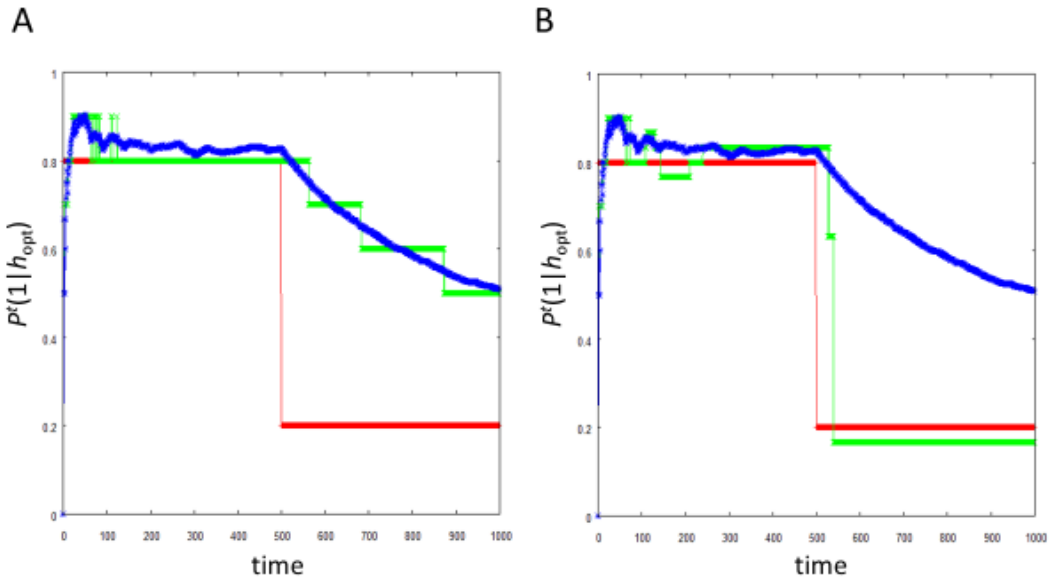

Figure 4

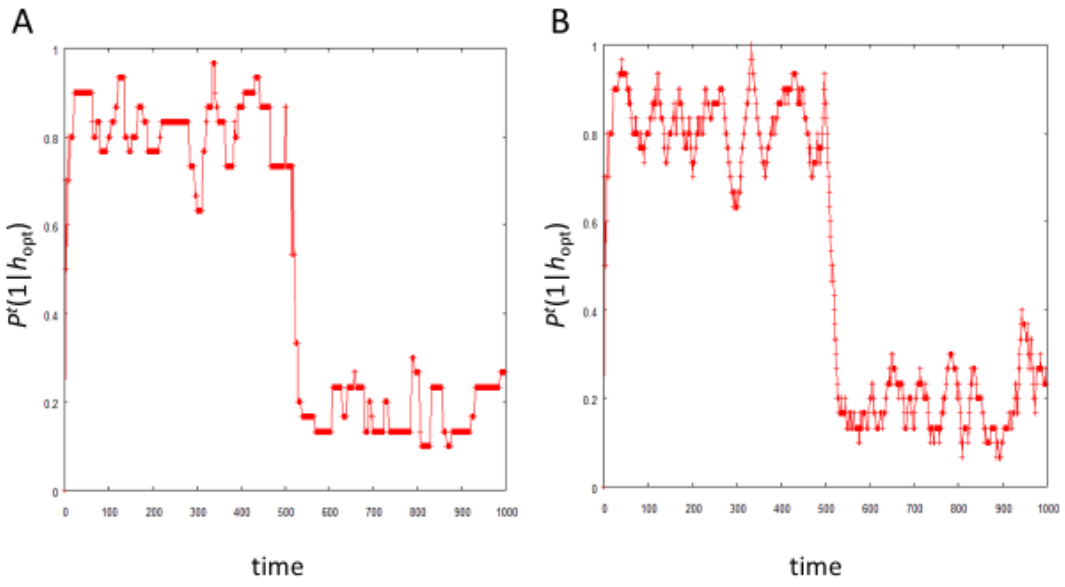
A
B
$P^t(1|h_{opt})$
time
0
0.2
0.4
0.6
0.8
1
0
100
200
300
400
500
600
700
800
900
1000

Figure 5

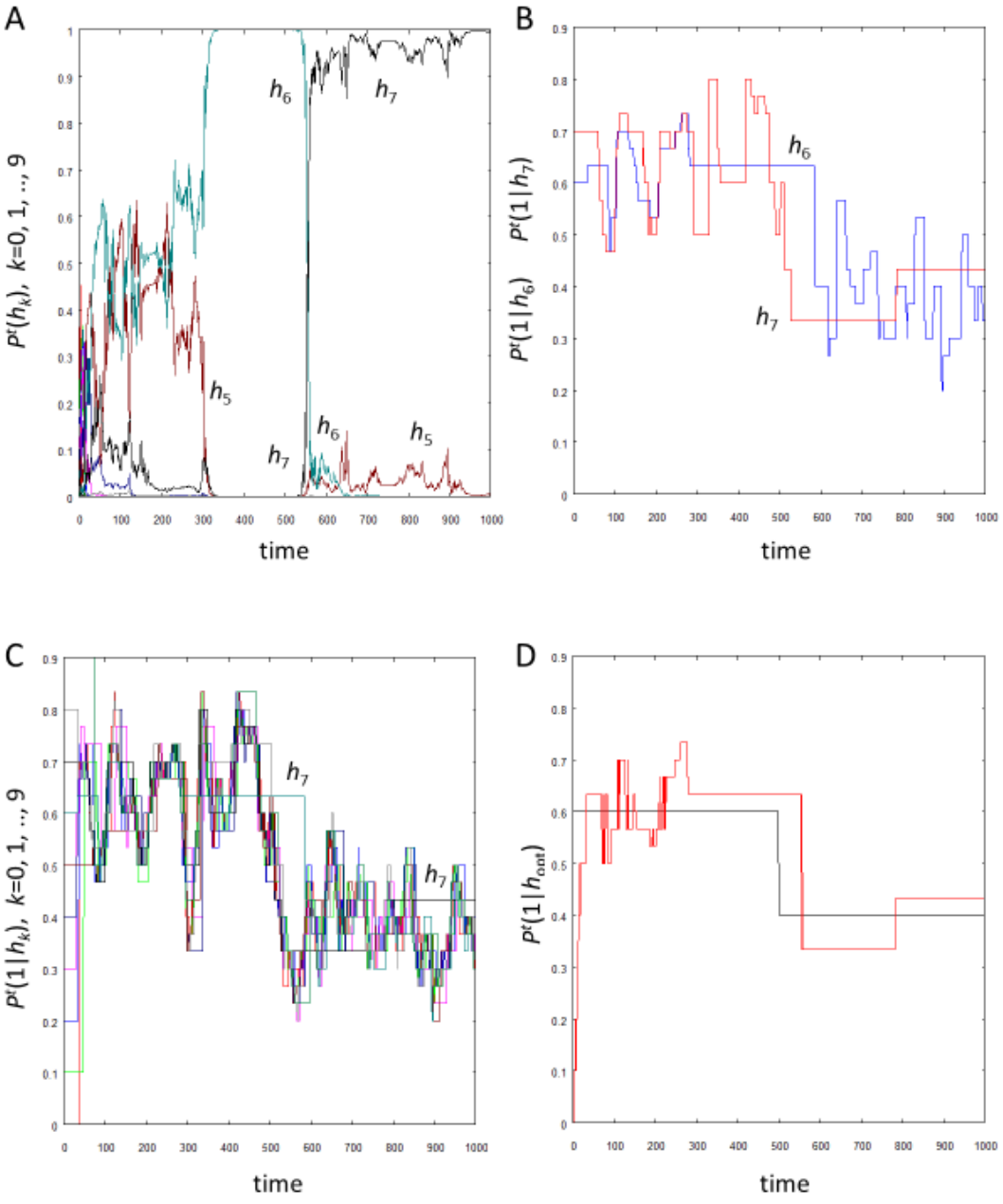
A
B
C
D
time
time
time
time
$P^t(h_k)$, k=0, 1, .., 9
$P^t(1|h_7)$
$P^t(1|h_6)$
$P^t(1|h_k)$, k=0, 1, .., 9
$P^t(1|h_{out})$
h_6
h_7
h_5

Figure 6

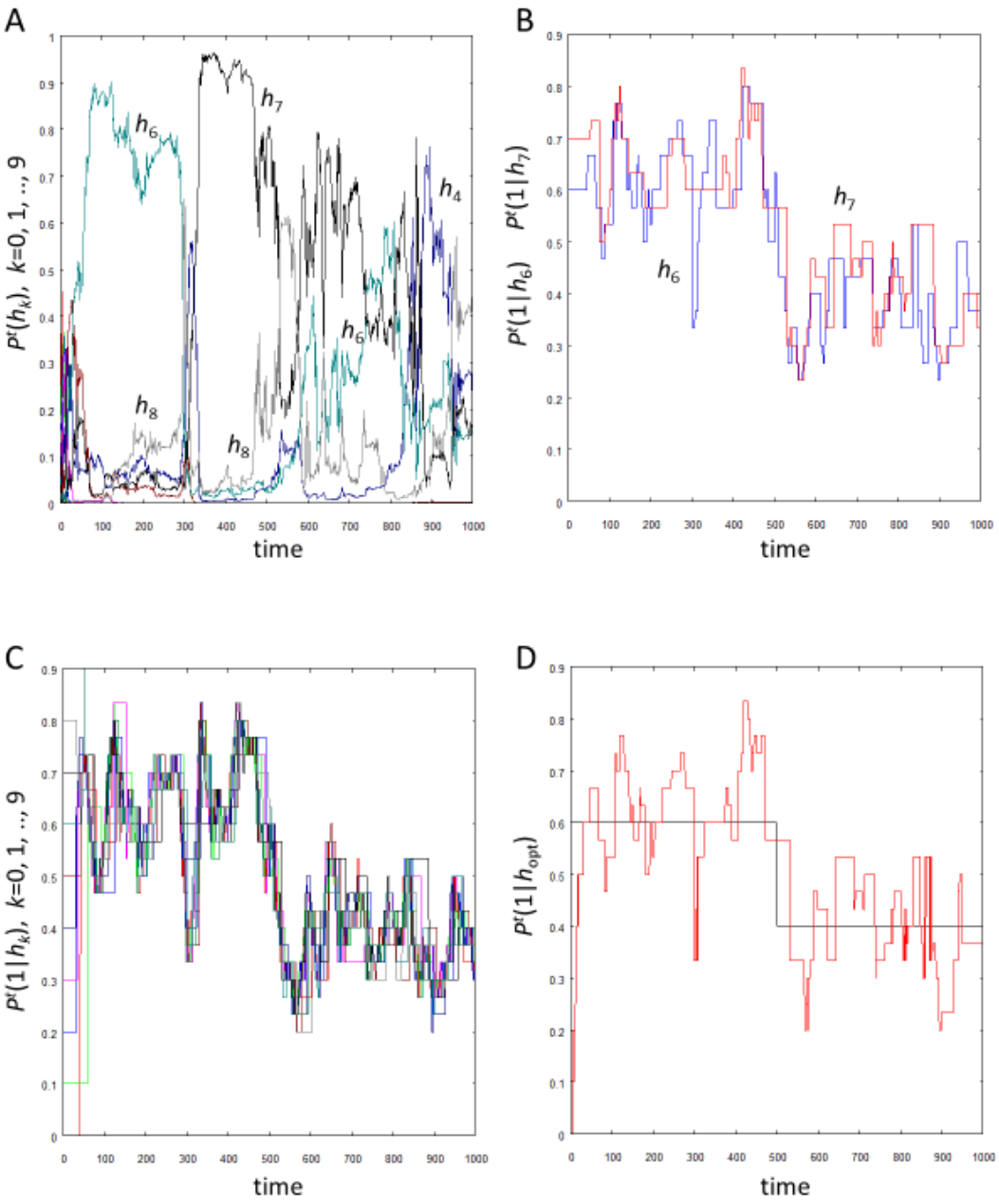
A
B
C
D
time
time
time
time
$P^t(h_k)$, k=0, 1, .., 9
$P^t(1|h_7)$ $P^t(1|h_6)$
$P^t(1|h_k)$, k=0, 1, .., 9
$P^t(1|h_{opt})$
h_7
h_6
h_4
h_8
h_8
h_6
h_7
h_6

Figure 7

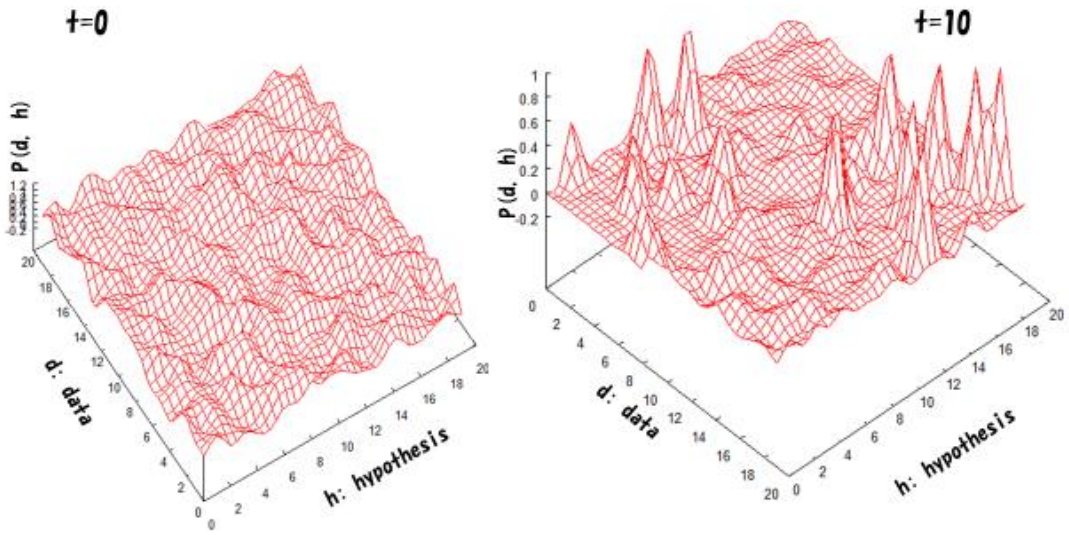

Figure 8

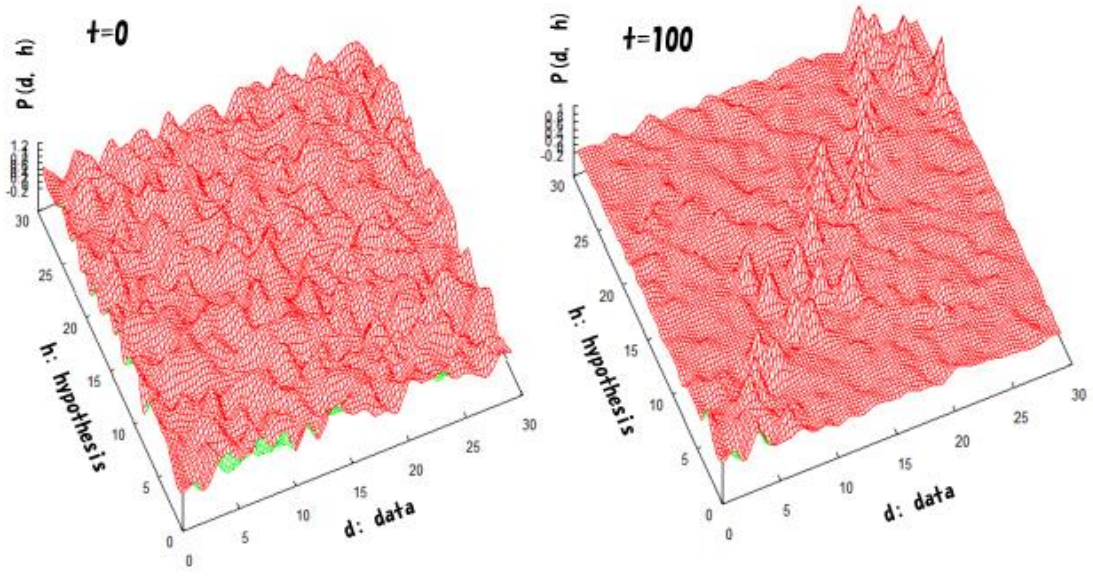

Figure 9

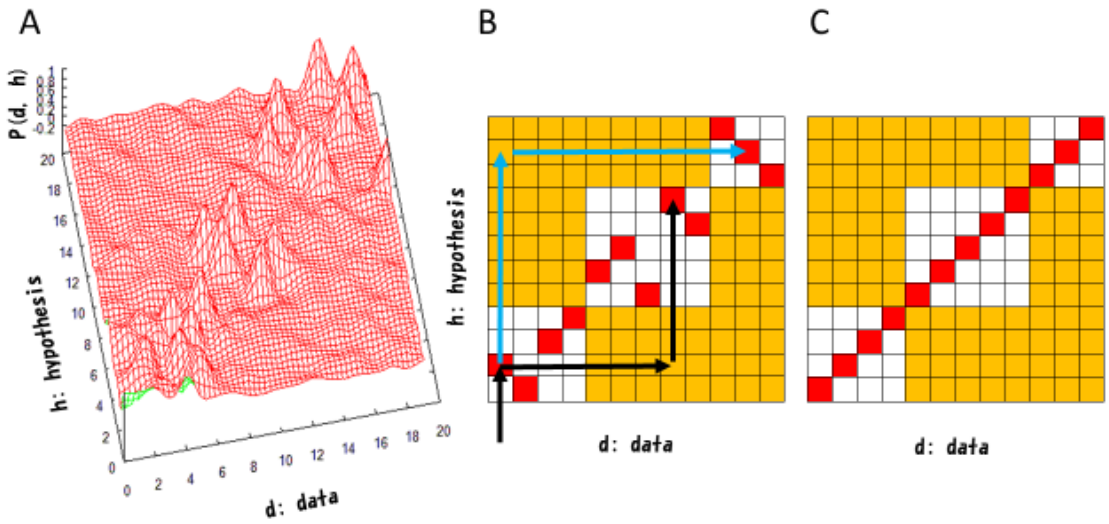

Figure 10

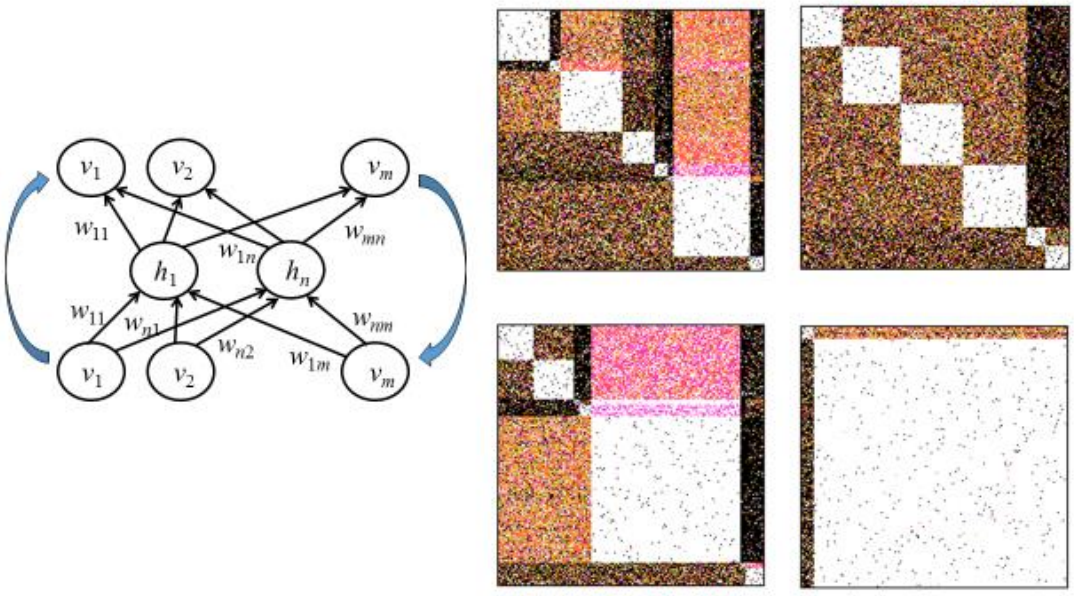

Figure 11

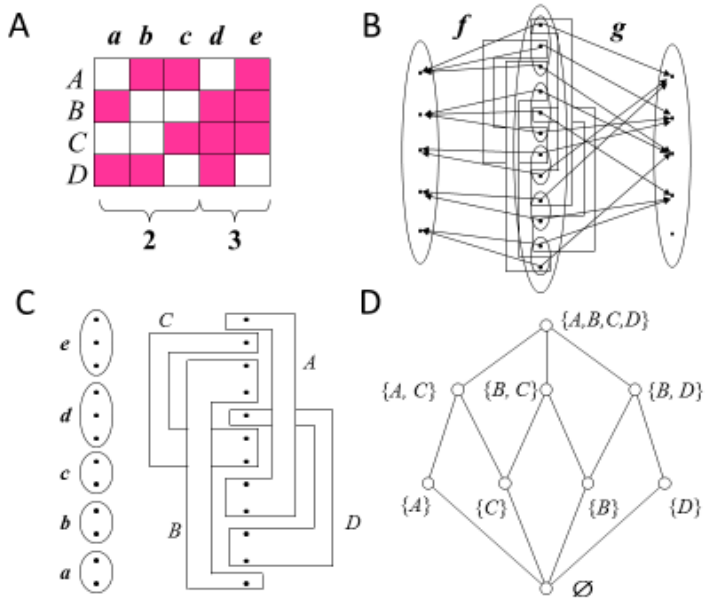

Figure 12

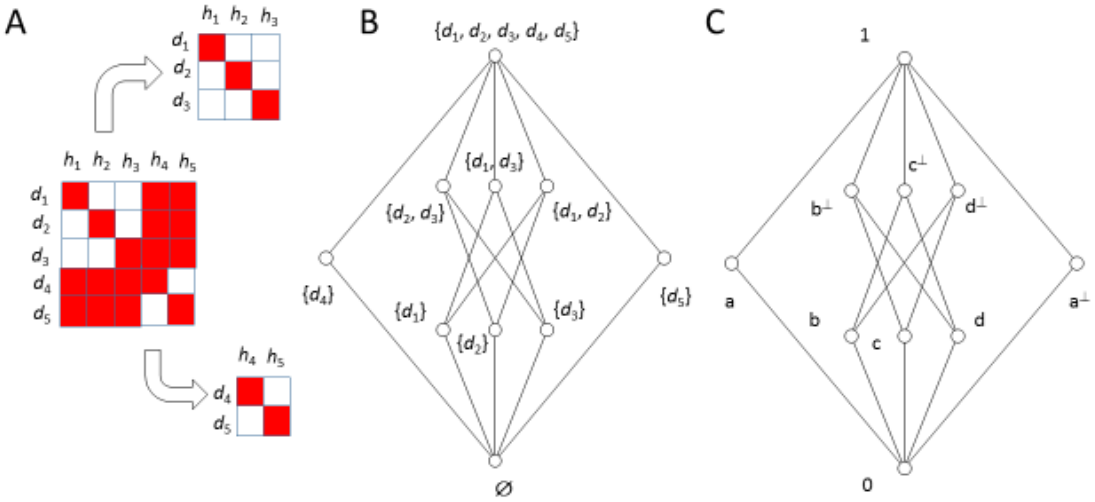

Figure 13

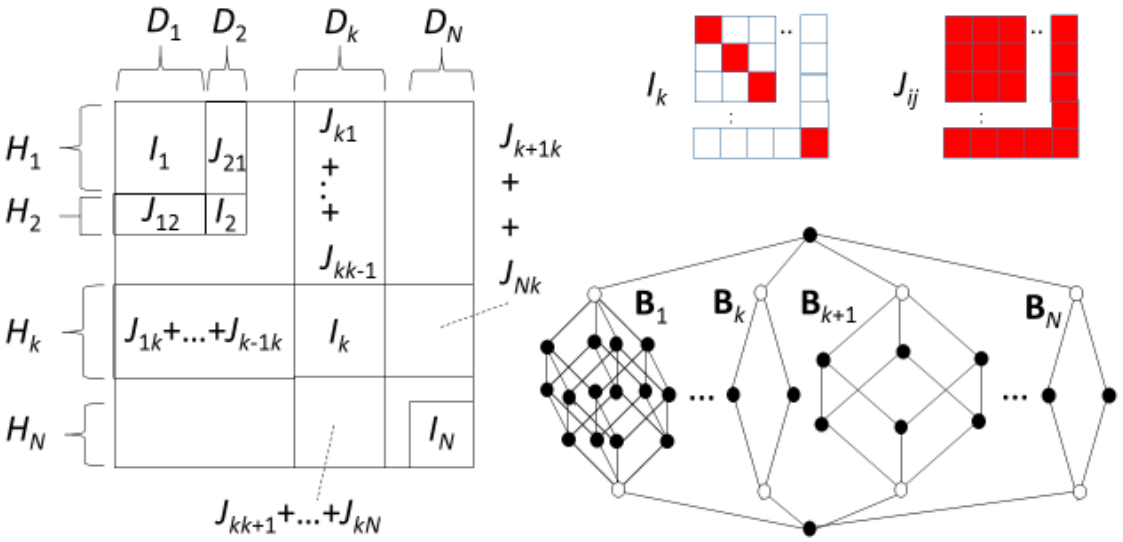

Figure 14

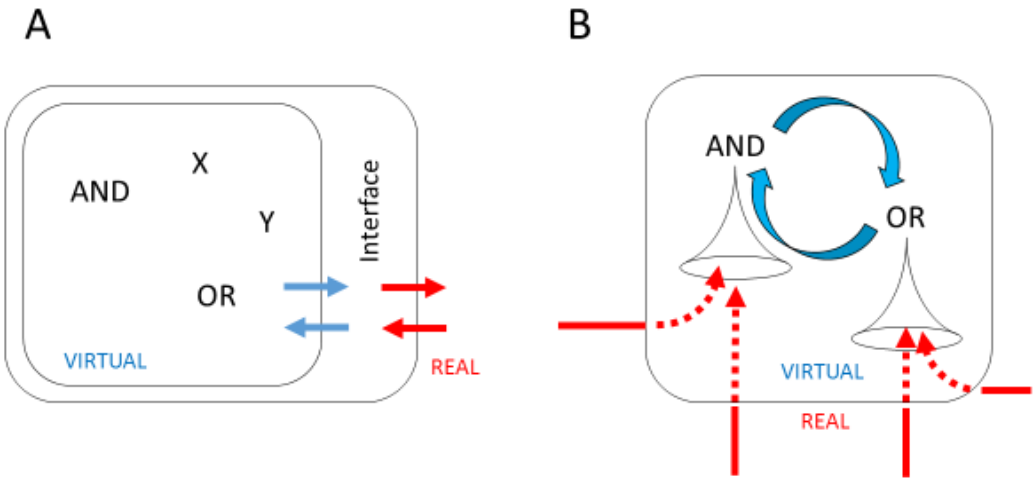

Figure 15