博士学位論文

論文題目: Time Scales, Synaptic Plasticity and Embodiment
(タイムスケール、シナプス可塑性、身体性認知)

松田　英子

Eiko Matsuda

東京大学 総合文化研究科広域科学専攻

Contents

List of Figures

6-2 **The process of learning.** A, At the beginning of each trial, the robot is placed on the central arm of the T-maze. The camera detects the black wall at the back, and the central part of its visual field is activated. B, The robot has not get learned much correlations about the tactile cue so that the robot makes no large turn. But a slight difference in the right and left motors, because of asymmetry from the training phase, leads the robot to one side of the T-maze. C, When the robot reaches the reward, the Hebbian learning rule reinforces the correlation between an increase of reward and the camera image of the black wall. D, After getting the reward, the robot is returned to the starting position to start the second trial. Through the "reward – vision – motor" path learned at the first trial, the virtual unit of the reward module (always set to 1.0) propagates into the vision module, inducing a left turn, i.e., a turn to the previous reward position. E, During the turning, the left whiskers perceive the tactile cue, so that a "tactile – motor" correlation is produced. 100

6-4 **Illustration of the learning process of the "tactile – motor" path in the second trial.** Five boxes in each module show activities of each of the five types of variables used for computation of Eq. 2 and 3. The sign in the boxes at the tactile module denotes reward(+), punishment(-) and nothing(0) for the left and right sides of the whiskers. A, Because of the "reward – vision – motor" path, the robot makes a left turn, and at the same time the left whiskers break contact with the tactile cue. Therefore, the robot learns a correlation between the decrease of the state unit $\Delta \boldsymbol{x}^{tactile}(t)$, and the state unit $\boldsymbol{x}^{motor}(t)$ that induces a left turn. B, After the second trial, if the robot touches the tactile cue on the left side, the state unit of the tactile module $\boldsymbol{x}^{tactile}(t)$ takes a positive value. The virtual unit of the tactile module $\tilde{\boldsymbol{x}}^{tactile}(t)$ is normally not activated because any correlations from the other modules to the tactile are not learned often. Therefore, the difference of the virtual and state unit $\Delta \tilde{\boldsymbol{x}}^{tactile}(t)$ takes a negative value. This negative activity propagates through the learned connections to induce a left turn. 102

List of Tables

Chapter 1

Introduction

Living systems are highly complex physical systems, which exhibit many unique characteristics in respect of self-reproduction, autonomy, enaction, sustainability and evolvability. The aim of this thesis is to approach a deeper understanding of living systems which exhibit amazing complexity and adaptivity, and to extract the essential logic underlying the intelligence of living systems.

A keyword of this thesis is “self-organization” of living organisms. Self-organization means a system where global patterns are spontaneously generated from many interacting sub-systems. For instance, similar patterns to fish skins are created from a simple reaction diffusion system, known as Turing patterns [56, 77]. Cell-like patterns known as Benard cells are spontaneously created by inhomogeneous distribution of temperature [87].

We assume that living organisms are also products of self-organization. They are organized through natural evolution, which does not have goal nor intention. Phenotypes and functions that can be observed in living systems are not specified in advance, but spontaneously organized through natural selection. However, the self-organized patterns presented so far, e.g., Turing patterns, may be able to imitate skins of living organisms, but far from living organisms themselves. In this thesis, we investigate general laws of self-organization underlying living organisms, i.e., self-organized structures that allow living systems to exhibit adaptive behavior.

Especially, we focus on self-organization of time scales. The ability to regulate time scales is a fundamental requirement for living systems. Sensory perception, cognitive processing and motor control all require precise control on temporal information, which is true to amebas to

humans [11, 12, 104]. Functions of living organisms work on various range of time scales: time scales in the range of tens of milliseconds are used for perception, e.g., sound localization, motion detection, or speech recognition [12]. Longer ones are observed in circadian rhythms, such as sleep-awake cycle or appetite [11]. The dynamical mechanisms for organizing those multiple time scales are still unclear. Indeed, it has not yet been determined how the well-regulated timescales are ubiquitously organized and what is the simplest unit to encode them.

The purpose of this thesis is to investigate 1) whether timescales can be self-organized in living systems, 2) whether having multiple time scales is efficient to generate adaptive behavior, and 3) how living organisms synthesize micro level time scales – or neuronal level and sensory / motor processing level – into macro level ones – or behavior.

The methodology of this thesis is firmly rooted in two bases: artificial life (ALife) and embodied cognition. Those two share a motivation to create a theory for understanding natural intelligence from physical and dynamical aspects. The philosophy of ALife lies in creating life-like phenomena from non-living matter, such as computer simulations, robots, or biochemicals. The main focus of this research field is not on mimicking minute details of living systems, but creating abstracted features of them, such as self-reproduction, autonomy, enaction, sustainability, and evolvability. This methodology has revealed essential principles underlying living organisms. Embodied cognition is a philosophy to understand the nature of intelligence. This theory suggest that natural intelligence is not designed by an external observer, but spontaneously emerges from complex interaction between the body, brain and environment, as a process of self-organization.

Based on the two bases, we explore self-organization processes observed in living organisms by creating two artificial systems: one is a robot, and the other is a system composed of biological neurons. The robot is equipped with a self-organized brain system, where we observe organization of adaptive behavior through a course of learning. There are several time scales concerning the robotic system: the fastest time scale lies in sensory processing, followed by learning speed, and the largest time scale concerns the generation of behavior. We observe coherent behavior, i.e., solving a T-maze, can be self-organized from the multiple time scale system. In the experiment with the biological neuron, we hypothesize that the material properties of the biological neuron innately has intelligent properties to organize multiple time scales. Biological neurons are cultured in *in vitro* and we record their spontaneous development with a micro-electrode

array. In both experiments, we observe and analyze self-organized structure of time scales, to finally synthesize them into understanding of living systems.

In the following sections, we first review the recent studies on time scales. Next, we detail basic philosophies underlying the thesis, i.e., ALife and embodied cognition, which is followed by general ideas and recent studies of self-organization. Finally, we present contribution of the thesis toward the theory of living systems.

1.1 Time scales

Many studies have suggested that regulation of time scales are important for maintaining various functions, as well as functions in turn work on different time scales. Living organisms have multiple time scales in the range from microseconds to days for realizing various functions [12]. The shortest time scale is observed in sound localization. Animals can localize a source of sound by discriminating a small difference in the arrival time of sound between the left and right ears [12, 14, 70]. The time scale – 10μs – is quite small compared to that of action potential which lasts some hundred microseconds. In a millisecond time scale, monosynaptic reflex arcs of neuronal cells are observed [66], i.e., one-to-one interaction of each single cell of neurons. It is on the scale from milliseconds to seconds that complex forms of sensory and motor processing occur, including speech generation / recognition [68,93] and motor control [11,27]. Time scales in the range from seconds to minutes concerns foraging [54], multiple-step arithmetic [100] and decision making [39,63,88,102]. Circadian rhythms display endogenous oscillations of 24 hours, which control the sleep-awake cycle, metabolic and appetite [20]. The longer ones than those are menses, menopause, and life span.

These studies have shown the importance of regulation of multiple time scales to maintain functions, as well as the fact that functions act on specific time scales. The work in this thesis concerns efficiency in having multiple time scales, which is introduced in the following part.

1.1.1 Efficiency of Having Multiple Time Scales

A study done by Dosenbach et al. suggests that multiple time scales in the brain allow robust control of behavior. By using fMRI meta-analysis techniques, they identified two distinct dominant

networks on top-down control, named “fronto-parietal” and “cingulo-opercular” networks [24]. They demonstrated that the two separate networks acted on different time scales. The front-parietal network was suggested to function on a short time scale, which allowed to rapid control on initiation and adjustment. On the other hand, the cingulo-opecular network appeared to work on a long time scale, which provided stable set-maintenance over a long time. They suggested that having several networks, which work on different time scales, allows flexible and robust control on human behavior.

1.1.2 Our Hypothesis

Based on the studies presented above, we hypothesize that 1) organization of the multiple time scales is an emergent property of the living organisms, 2) the multiple time scales are efficient for creating robust and adaptive behavior, and 3) living organisms synthesize the micro level time scales – or neuronal level and sensory / motor processing level – to create macro level time scales – or adaptive behavior. We investigate 1) –3) by a robotic model in Chapter 2, and 1) is intensively studied in Chapter 3 by characterizing a developmental process of biological neurons *in vitro*. In the next sections, we present core concepts that underlie our experimental setups.

1.2 ALife

The central question of ALife is how to make life from non-living matter. The methodology of ALife is to construct abstract models of life-like phenomena by using artificial systems, such as computer models, robots, and biochemistry, which enables to extract essential logic underlying living organisms. Main topics of ALife lie in self-reproduction, autonomy, enaction, sustainability, evolvability, and so on [8, 44, 76, 83, 86]. Namely, the main focus of ALife is not on creating elaborate models of living systems, but to realize essential features of them. The researchers have been aiming at revealing “life-as-it-could-be” rather than “life as-we-know-it”; the broader sense of life tells the essential logic underlying living organisms [62].

A series of models that we present in this thesis is based on the philosophy of ALife. The models are not biologically elaborate, but simplified abstract ones. Those models, we believe, reveal essential properties that underlie living organisms. In this section, we first review the

history of ALife, and then introduce relevant backgrounds with the works of this thesis.

The field of ALife was founded by a computer scientist, Christopher Langton [62]. He found out that self-reproduction – one of the essential properties of living systems – can spontaneously emerge from a set of simple algorithms in a computer [61]. He concluded that, by giving simple rules like algorithms, life-like phenomena can be generated, which would contribute to understanding essential principles of living systems.

An earlier example of ALife is the cellular automaton [80,110,114,117]. A cellular automaton is composed of a grid of cells, where each cell can take a finite number of states, e.g., 0 and 1 [117]. The state of each cell is temporary updated, depending on the neighboring cell states. Depending both on the rule to update the state and the initial configuration, a cellular automaton can exhibit complex patterns, e.g., similar patterns to surfaces of sea shells [80, 117], and self-moving patterns [1,40,42]. Autonomy is one of the essential properties of life, so that the cellar automaton is one of the most famous examples in ALife.

Models using biochemicals have also been presented. For instance, Hanczyc et al. demonstrated that oil droplets composed of fatty acids exhibited autonomous movement [43]. The movement was driven by the internal convection flow. The convection was induced by a hydrolyzation process of anhydride precursors of the fatty acid, which happened on the droplet surface. The movement of the oil droplet was sustainable, because of a positive feedback to allow fresh precursors to react. This experiment suggests that life-like phenomena like autonomous movement can be self-organized only from material properties.

In Chapter 3, we investigated self-organized structures observed in biological neurons *in vitro*. We hypothesized that material properties of neurons can innately generate life-like phenomena. Material properties of neurons mean: 1) components such as neural somata, glias, axons, dendrites, synapses, 2) physical restrictions like transmission speed of electrical signals, thermal noise or nutrition supply and 3) plasticity of synapses that allowed development. Artificial neural systems that have been used in ALife studies often ignore those properties of neurons. We hypothesize that neurons innately have intelligent properties such as controlling multiple time scales, so that we investigate self-organized patterns observed in the biological neurons.

1.3 Embodiment

Embodied cognition is a theory to understand the nature of the mind, which is oriented toward building intelligent behavior artificially from non-living matters, and therefore studied mainly in the field of ALife and artificial intelligence [4,5,7,9,10,18,21,23,26,48,60,64,76,86,107,113]. In this section, we first review the history of embodied cognition, and then mention the contribution of the works in this thesis.

Several decades ago, the field of artificial intelligence assumed the *computational* view of the mind. In this approach, the brain was regarded as a central controlling system, which was distinct from the body and sensors. The brain received information from sensors, processed it to choose a good plan from given behavioral patterns, and changed the motor output to generate the desired behavior. Namely, the brain worked as a search system to pick up a good plan. To do so, the brain system needed to have a perfect representation of the environment, e.g., maps, locations of objects, or well-organized tasks. The slogan there was "good representation is the key to AI" [8]. The researches in the field hoped that intelligence could be achieved by enough knowledge of the environment and complex computation to manipulate given symbols.

However, it has been realized that the computational approach has a lot of suspicious assumptions and qualitative differences from living systems. Because of the need for a perfect representation of the world, the robots needed heavy computation, which after all prevented them from exhibiting naturally smooth movement. Additionally, they were quite fragile to unexpected perturbations, because it was actually impossible to obtain the complete representation in advance.

Brooks presented a series of robotic experiments that changed radically the philosophy behind robot control. His robots were composed of several parallel control systems, e.g., avoiding objects, and they exhibited surprisingly complex and robust behavior like exploring in a rough topography [8]. He suggested that the central nervous system and perfect representation were not necessary to create robust and intelligent behavior; Accumulation of basic functions that were controlled by small subsystems can finally be synthesized into complex, and observably "intelligent" behavior. Namely, intelligent behavior is not designed in advance, but self-organized through interaction between the body, brain and environment. Beginning from his idea, a new

scientific field has grown, known as embodied cognitive science or embodied artificial intelligence.

Embodied cognition suggests that the nature of the mind is determined by the physical characteristics of the body, brain and the environment [8,83,86], rather than controlled by the central nervous system. This notion does not simply means that intelligence requires body, but intelligence comes from complex interaction between the body, brain, and environment. Namely, intelligence is dependent on 1) morphology and material property of the body and 2) the physical interaction with the surrounding environment [84].

For instance, Pfeifer presented a robotic experiment to show that intelligent behavior comes from morphology of the body [85]. The robot had a task to collect garbage cans, that were scattered in the environment. It should not collect larger objects than the cans, because they were not garbage. The robot had a gripper on its body, which was just fit to the cans. When the robot moved around the environment, it successfully collected the cans, by catching them with the gripper. It ignored other objects, because the robot could not catch larger objects because the gripper was too small to do it. This example showed that robot was able to discriminate garbage from other surrounding objects only by the morphological structure of the body, which finally generated garbage collecting behavior [85]. Namely, he showed that intelligent behavior is generated from the morphological structure of the body and interaction with the environment

In the research field of embodied cognition, it is quite important to construct a physical system in practice before creating a theory. Creating actual systems allow to realize essential characteristics to build a theory which otherwise cannot be realized. This type of approach is called “constructive approach”. The slogan of this approach is “understanding by building” [83]. This approach not only allows to dismiss wrong assumptions, but also to identify general principles underlying the physical system that are hard to predict before creating the system in practice. Bovet metaphorically illustrated this point in [6] through considering the history of the question about how birds can fly: Once people assumed that the essential mechanism for birds to fly was the presence of wings. They did not realize the assumption was actually not true until they constructed artificial flying systems, where they first realized the importance of other factors like weight distribution. Building artificial systems allows evaluation of improper assumptions, and reveals general principles that otherwise cannot be predicted, which are finally synthesized

into a theory, aerodynamics for example.

The concept of embodied cognition is one basis of this thesis as well as the concept of ALife. We aimed at understanding adaptive behavior by building artificial systems which interact with the environment by using the physical bodies. We build a robotic experiment in Chapter 2, where the robot is equipped with a self-organized brain system and multiple sensors and motors. The robot is controlled by hierarchical orders of multiple time scales, i.e., time scales to control 1) neuronal activity (fastest), 2) synaptic connectivity, and 3) behavior (longest). We analyze how the faster time scales are organized into behavior. Furthermore, we analyze time scales of information flow in the neural network, and discuss the efficiency of having multiple time scales for adaptive behavior.

1.4 Self-organization

As presented earlier, we investigate self-organization of adaptive behavior observed in two different systems: Robotic experiment and biological neurons *in vitro*, both of which have learning capability to demonstrate self-organization through learning / development. The robot has self-organized brain system based on Hebbian learning, while the biological neurons have synaptic plasticity to evolve through the culturing. In this section, we introduce the background of self-organization processes observed both in the robotic experiment and the biological neurons.

1.4.1 Self-organization in robotic systems

First, we introduce a history of self-organization studied in the field of evolutional robotics. Braitenberg presented “Braightenberg vehicles” in a series of thought experiments, which demonstrated self-organized behavior in simple robots [7]. Simple wiring between sensory and motor systems can spontaneously generate surprisingly complex and intelligent behavior, e.g., light seeking behavior. As recent studies, Kuniyoshi et al., presented the emergence of intelligent behavior without having motivational principles like an evaluation function [57–59]. Their computational models were equipped with muscle and skeleton systems and controlled by chaotic oscillators. The robots spontaneously generated meaningful and adaptive behaviors through moving around the environment by using the physical bodies. Their experiments suggest that

coherent behavior can be self-organized through interaction between the environment and the physical body, where there is no need for explicit evaluation functions.

More recently, Bovet and Pfeifer invented a robot with a self-organized brain system [4–6]. In their model, the robot was equipped with motors and several types of sensors, which were controlled by corresponding sensory or motor areas in the brain. In their self-organized brain system, Hebbian learning rule, which captures simultaneous activities of neurons [45], tuned connectivity strength between the sensory / motor areas. The robot did not have any evaluation functions, but only by interaction with the environment and their body, brain and sensors, they exhibited meaningful behaviors such as avoiding walls or solving delayed reward tasks.

The robotic model that this thesis presents in Chapter 2 is based on the model by Bovet and Pfeifer [4–6]. We analyze behavioral change during a course of learning, where we expect to observe similar learning patterns to living organisms. We further analyze the self-structured learning patterns, adaptivity of them, and neuronal mechanisms to create the behavior. We especially focus on self-organized structure of time scales to allow adaptive behavior.

1.4.2 Self-organized structure in biological neural networks

Self-organization observed in neuronal systems is tightly related with synaptic plasticity, which is the ability to tune synaptic strengths in response to activities of neurons. A basic mechanism of the synaptic plasticity is Hebbian learning rule [45]. This learning rule concerns adaptation of neurons through learning, where synaptic strength increases if the postsynaptic cell is activated just after the activation of the presynaptic cell. Hebbian learning is experimentally observed in real brain systems as spike-timing-dependent-plasticity (STDP) [3,13,13,15,28,31,37,41,99,101, 109].

However, it is yet to be fully investigated whether STDP neural networks can generate computational functions, and if so, how. Nessler et al. are the precursors in answering the questions; they showed, through a computational brain model, that Bayesian computation emerges in the neural network equipped with STDP [75]. They further demonstrated that STDP approximates the most powerful learning principle, Expectation-Maximization (EM) [22]. This theoretical study suggests that even small number of neurons can exhibit computational functions.

In Chapter 3, we study a self-organization process of neural networks cultivated *in vitro*,

where we examine emergent properties that the biological neuron naturally possesses, especially focusing on time scales.

1.5 Methodology

In this section, we would like to introduce three basics of our methodologies used in Chapter 2 and 3. The first two, i.e., genetic algorithm and learning, are important in generating self-organized structures, and the last one, transfer entropy, is to characterize self-organized structure in the neural networks.

1.5.1 Genetic algorithm

Genetic algorithm (GA) is popular methodology used in ALife researches, which mimics the process of natural selection in computer simulations [69]. GA allows virtual creatures in the simulation environment to evolve adaptive phenotypes, without specifying desired phenotypes in advance. Various unexpected phenotypes can be self-organized, similarly to the natural selection. Because of its capability to show a self-organizing process, we applied GA to our robotic model to tune parameters. Here, we first review the basic notion of GA, and then application of it.

The simulation of GA has a population of artificial creatures and a fitness function to evaluate behavior of respective creatures. Each creature has a set of phenotypes (defined by a genome), which is modified gradually through the simulated evolution, by giving mutations and recombinations to the genome. Depending on the phenotypes, each creature moves around the simulated environment, where the performance of each creature is evaluated by the fitness function. The fitness value determines the reproduction rate, so that creatures that exhibit better fitness values become evolved. In other words, GA is a method to search local optima in the fitness function.

Based on GA, Sims created evolutional creatures in a computer simulation, where the fitness function evaluated their capability to forage and escape from enemies [98]. The artificial creatures exhibited unexpectedly various strategies, where they could move, swim or jump depending on their evolved morphology. The variation of strategies comes from a lot of local optima in the fitness function. As such, GA allows to observe “emergent” solutions that otherwise cannot be

predicted. This evolutional principle is similar to the natural selection, in respect that there are no outsiders to arbitrary design each creatures, but adaptive behavior spontaneously emerges through the evolution. GA is used to evolve various life-like phenomena, from arm movements, decision making process, to eco-systems [74, 78, 98].

1.5.2 Learning

One striking difference between the living system and the evolved creature by GA lies in learning capability. Normally, the creatures evolved by GA has fixed parameter sets, so that they do not learn from ongoing interactions with the environment. However, learning is essential property of intelligent behavior of living organisms, because the ongoing control will allow robust and flexible control of behavior. In this part, we introduce the importance of learning in adaptive behavior by taking an example of a rat experiment.

In a route selection task, rats shook their heads at the intersection point, as if they were considering their way. This behavior is called vicarious trial-and-error (VTE) [108]. VTE was observed when rats were uncertain in making proper decisions, e.g., in an earlier stage of learning, error trials, the next trial after making an error, or when the rats had to change their strategy [94, 108]. Therefore, VTE is regarded as being relevant with learning efficiency. Place cell activity in rats' hippocampus revealed that VTE was observed when the place cell activity swept forward from the current position of the rats [52]. This suggests that they were simulating their next decisions internally under the uncertain situation. This leads to a conclusion that VTE reflects deliberative decision-making, which is a cognitive process that includes searching, predicting, and evaluating future outcomes [111]. These studies show that uncertainty during learning is critical to evaluate and determine their way. This process will be computationally slow but also adaptive because it allows ongoing control to achieve flexible behavior.

The robotic experiment presented in Chapter 2 focuses on self-organized behavior through learning, where we demonstrate spontaneous generation of VTE, and robustness allowed by it. We further investigate neural mechanisms to generate VTE in respect of organizing multiple time scales. In the next part, we introduce the way to assess time scales spontaneously organized in the neural network.

1.5.3 Transfer entropy

We apply transfer entropy (TE) to measure information flow in neural networks both in the robotic model (Chapter 2) and in the biological neurons (Chapter 3). TE is a measurement of directed information flow, or more specifically, it detects causal relationship between two time series [2, 65, 95, 103]. High TE in a directed pair of neurons means that high amount of information is transferred there. By using TE, we estimated information flow in different time scales.

Computation for TE needs to binalize raw data into 0 and 1, which needs to set a bin-length: If there are any neural spikes within a bin, then the bin will take the state 1, and otherwise 0. We took the bin-length as an indicator of time scales. If the bin-length is small, the binalized time series represents fast activity of neurons, e.g., single cell interactions. Larger bin-lengths capture macroscopic activity created by ensembles of neural activities. By changing the bin-length, we quantify information flow in multiple time scales. We detail the way to compute TE in the following part.

Background and calculation procedure of TE

Shannon 1948 founded the information theory [97], by defining a quantity to measure the amount of information. The quantity is called Shannon entropy. We first give an intuitive understanding of Shannon entropy, and specific definition follows next.

The Shannon entropy quantifies the amount of uncertainty in a random variable. As an example, consider coin tosses which are completely fair. Each trial happens independently, so that the history of the outcomes does not tell any information to lessen the uncertainty of the next outcome, which gives high entropy. On the other hand, Markov process, where the probability of the current outcome is dependent on a history of the outcomes, has lower entropy than the coin tosses, because the history lessens the uncertainty to give better prediction. Namely, the entropy is a quantity to measure the amount of uncertainty in a variable.

Specifically, Shannon entropy of a system X is defined as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x), \tag{1.1}$$

where $p(x)$ denotes the probability of x (x is an event of X).

To evaluate the dependency between system X and Y, mutual information (MI) is defined as follows:

$$\begin{aligned} MI(X,Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X). \end{aligned} \tag{1.2}$$

$H(X|Y)$ means conditional entropy, i.e., the uncertainty of X when Y is known. Therefore, $MI(X,Y)$ suggests a decrease in the uncertainty of X when Y is known.

TE measures the causal relationship between X and Y by considering the past history. The TE from system X to Y is denoted by $T_{X\to Y}$, which is written as:

$$T_{X\to Y} = H(y_{n+1}|y_n^{(k)}) - H(y_{n+1}|y_n^{(k)}, x_n^{(l)}) \tag{1.3}$$

$$= \sum_{n=0}^{N-1} p(y_{n+1}, y_n^{(k)}, x_n^{(l)}) \log \frac{p(y_{n+1}|y_n^{(k)}, x_n^{(l)})}{p(y_{n+1}|y_n^{(k)})} \tag{1.4}$$

where n is the current time step, and $y_t^{(k)}$ and $x_t^{(l)}$ are the past variables with length k and l respectively (i.e., $y_t^{(k)} = \{y_n, y_{n-1}, ...y_{n-k+1}\}$ and $x_n^{(l)} = \{x_n, x_{n-1}, ...x_{n-k+1}\}$). When the next state of Y ($= y_{n+1}$) is conditioned by the past history of X ($= x_n^{(l)}$), then $H(y_{n+1}|y_n^{(k)}, x_n^{(l)})$ takes a smaller value than $H(y_{n+1}|y_n^{(k)})$. If y_{n+1} is independently determined from the past history of X, then the two components in Eq. 1.3 will have the same value. Therefore, higher $TE_{X\to Y}$ means higher causality from X to Y.

1.6 Outline and Contributions

The following Chapter 2 presents a robotic experiment for understanding decision making processes spontaneously generated from Hebbian learning [4–6]. The behavior is self-organized through interaction with the body, brain and surrounding environment. We obtained 16 robots by GA that were successful in a given task, and found some of the robots showed conflict-like behavior. The robot shook its head direction more often at the beginning of learning, which decreased afterward. Similar behavior was found in rat's experiments, i.e., vicarious trial-and-error

(VTE; see Section 1.5.2). The result suggests that VTE is an emergent property of the system, generated from a simple interaction between the body, brain and environment. By analyzing the VTE from the chaos theory and the information theory (introduced in Section 3.3.3), we studied the role and mechanism of it. Specifically, we found that VTE-showing robot had multiple time scales in the brain system, and showed robust control toward perturbations. This demonstrates an efficiency of having multiple time scales toward adaptive behavior.

Chapter 3 investigates a intrinsic property of biological neurons. We cultured neurons *in vitro* to recode activities of neurons during several weeks. We observed a developmental process of the neurons, and characterize information flow by TE. We revealed that even small neural networks composed of roughly 100 neurons spontaneously develop multiple time scales in information flow.

Finally, a summarizing discussion of the general principles drawn from the results obtained throughout the thesis is provided in the concluding Chapter 4.

1.6.1 Major contributions

The major contributions of the thesis are summarized as follows:

- We demonstrate that similar behavior to the rat (i.e., VTE) can be self-organized in a robotic model with hierarchical time scales, i.e., sensory processing, learning, and behavior, presented in Chapter 2.

- We evaluate the hypothesis that multiple time scales in the neural network generates adaptive behavior, and are efficient in robustness, proposed in Chapter 2.

- The result in Chapter 3 shows that multiple time scales are an emergent property of biological neurons *in vitro*.

Chapter 2

A Robotic Approach to Understanding the Role and the Mechanism of Vicarious Trial-and-Error in a T-Maze Task

Vicarious trial-and-error (VTE) is a behavior observed in experiments with rats that seems to suggest self-conflict. This behavior is seen mainly when the rats are uncertain about making a decision. The presence of VTE is regarded as an indicator of a deliberative decision-making process, i.e., searching, predicting and evaluating outcomes. This process is slower than automated decision-making, such as reflex or habituation, but it allows for flexible and ongoing control of behavior. In this study, we propose for the first time a robotic model of VTE, to see if VTE can emerge just from a body-environment interaction, and to show the underlying mechanism responsible for the observation of VTE and the advantages provided by it. We tried several robots with different parameters, and we have found that they showed three different types of VTE: high numbers of VTE at the beginning of learning, decreasing afterward (similar VTE pattern to experiments with rats); low during the whole learning period; and high all the time. Therefore, we were able to reproduce the phenomenon of VTE in a model robot using only a simple dynamical neural network with Hebbian learning, which suggests that VTE is an emergent property of a plastic and embodied neural network. From a comparison of the three types of VTE, we demonstrated that 1) VTE is associated with chaotic activity of neurons in our model and 2) VTE-showing robots were robust to environmental perturbations. We suggest that

the instability of neuronal activity found in VTE allows ongoing learning to rebuild their strategy continuously, which creates robust behavior. Based on these results, we suggest that VTE is caused by a similar mechanism in biology, and leads to robust decision-making in an analogous way.

2.1 Introduction

In a study with rats, Tolman [108] observed that they seemingly hesitated when they had to choose between one of two rooms, one of which contained a reward while the other was empty. The only cue differentiating the rooms was the color of their doors. A black door indicated that the room provided a reward, and a white door indicated an empty room. To reach the reward, the rats had to learn the relationship between the color of the door and the presence of the reward. During the learning phase, the rats were seen moving their heads from one door to the other as if they were considering which one to choose, which was referred to by Tolman as a conflict-like behavior called vicarious trial-and-error (VTE). In his experiments, Tolman noticed that the number of VTE events (i.e., the number of times that the rat shaked its head during one trial) increased at the onset of the learning phase but started to decrease when the performance was stabilized. Based on this observation, VTE has been connected to learning efficiency.

Following Tolman's observations, other researchers started paying attention to the presence of VTE. Hu and Amsel showed a hippocampal contribution to VTE [47]. Johnson and Redish recorded place cell activity in rats' hippocampus, and they observed VTE when the rats were simulating their next decisions internally before acting [52]. These results led to the hypothesis that VTE reflects deliberative decision-making, which is a cognitive process that includes searching, predicting, and evaluating future outcomes [111]. This process is computationally slow compared with automated decision-making like habituation and reflex, but it allows ongoing control to achieve flexible behavior. A rat experiment by [94] supports this hypothesis; the author observed high VTE when the rats were uncertain and had to think about their decisions in the following three conditions: 1) error trials, rather than correct trials, 2) the next trial after making an error (i.e., potential error trials), and 3) when the rats had to switch their strategy.

VTE-like behavior has also been found in other animals. In a human experiment [115], the

participants showed VTE-like behavior when they had to actively, instead of passively, explore a given environment. They performed better with VTE-like behavior in the active condition. Tarsitano et al. [106] found that in a detour task, jumping spiders displayed two phases of action: the inspection phase, when the spiders stopped and inspected possible routes toward a target, and the locomotory phase, when the spiders moved toward a single direction. VTE was observed during the inspection phase. Tarsitano concluded that "one can speculate that it is a small but significant jump to use trial and error when choosing a goal to approach". However, in animal experiments it is difficult to observe the neural dynamics, which makes it hard to directly investigate the mechanism behind VTE and its resulting behavior.

There are some theoretical models of VTE; Rossler took up VTE as a sign of private simulation [90], and Ikegami studied it as an example of embodied chaotic itinerancy [50]; i.e., the itinerant motion of an autonomous robot with chaotic instability. From the Bayesian theory view point, Johnson et al. posited that VTE occurs with changing task demands [53]. Still, those models are lacking body and environment structure that will spontaneously generate VTE.

In this paper, we tried to understand whether VTE is an emergent property of a physical body moving around in its environment. We also investigated the link between VTE, neuronal dynamics, and efficiency of VTE toward learning. The methodology used in this chapter is to make a simplified abstract model of VTE, rather than making a biologically elaborate one. This experiment has its basis in the field of evolutionary robotics, where basic features of living organisms are recreated by simple robotic systems [44, 76]. This is to study the essential logic underlying living systems, such as autonomy, evolvability, and embodiment [8, 44, 76, 83, 86]. In this paper, we especially focused on embodied properties of living organisms. Embodied cognition is the notion that the nature of the mind is determined by the physical characteristics of the body and the environment, rather than controlled only by the central nervous system [8, 83, 86]. For instance, Bovet and Pfeifer [5] showed the spontaneous development of coherent behaviors in robotic experiments – reward seeking behavior for instance – just by moving around the environment with a physical body. In their experiment, the robot utilised the morphological structure of whiskers and the physical properties of various sensors like infrared or vision sensors. By learning the relationship between those different types of signals, the robot successfully reached the reward. This experiment gives an important notion that intelligent behavior can emerge from the simple

interaction between the body and the environment.

We therefore composed a simple robotic model of VTE to show the underlying mechanism responsible for the observation of VTE and the advantages provided by it, in terms of learning and in terms of dynamical systems. The model that we used is based on Bovet's T-maze learning robots [5] which we reproduced in computer simulations. The model that we used is similar to the environment used when the rats showed VTE. As a result, we demonstrated analogous VTE patterns to those reported in the experiment with rats, in terms of the temporal change in the number of observations, i.e., high at the beginning of learning and less afterward.

In addition to the similar pattern to rats, we have found other patterns of VTE using different parameters, which we classified into the following three groups: 1) high number at the beginning of learning and low afterward, 2) low during the whole learning period, 3) high all the time. From the comparison of the types of VTE, we demonstrated that VTE is associated with the chaotic activity of neuronal dynamics. Depending on the three types of VTE, we also compared the robots' behavior to evaluate adaptability by changing environmental conditions. The robots with the low VTE pattern changed their behavior drastically due to those perturbations, even exhibiting 0% success, while those with the high to low VTE pattern were robust to the perturbations. we suppose that VTE causes sensory fluctuations, enabling the robot to continuously change the connectivity pattern of the neurons. In other words, VTE allows the robot not to have to follow the same trail in a maze, enabling it to change its neural activity. This allows the robot to learn in an ongoing way, by continuously gathering information from the environment, creating robust behavior. This reminds us of the concept of "homeodynamic adaptation" suggested by Iizuka and Di Paolo [49] – an agent-based cognitive model of morphological disruption where internal instability allows behavior that is adaptive to changes in the body.

The organization of this paper is as follows: In Section 2.2, we detail the environmental set-up and the neural model. Section 2.3 provides results and analysis of the behavior of the robots. Section 2.4 discusses the role of VTE and the mechanism behind it in the light of adaptability, which are concluded in Section 2.5.

2.2 Methods

This work is based on a robot experiment developed by Bovet and Pfeifer [5]. In this model, a robot must reach a goal located on one arm of a T-maze. To do so, a neural network acting as the controller must combine four sensory modules to determine the right motor commands at every instant (Figure 2-1A). The sensors are tactile, visual, proximity, and reward. These five modules (i.e., the four sensory modules and the motor module) are interconnected and represented by a different neural population within the controller. Synaptic connections are tuned using the Hebbian learning rule, which will be explained in more detail later on. Differently from the original work by Bovet and Pfeifer, the experiments here were done in computer simulation. The robot was modelled according to the e-puck robot [71].

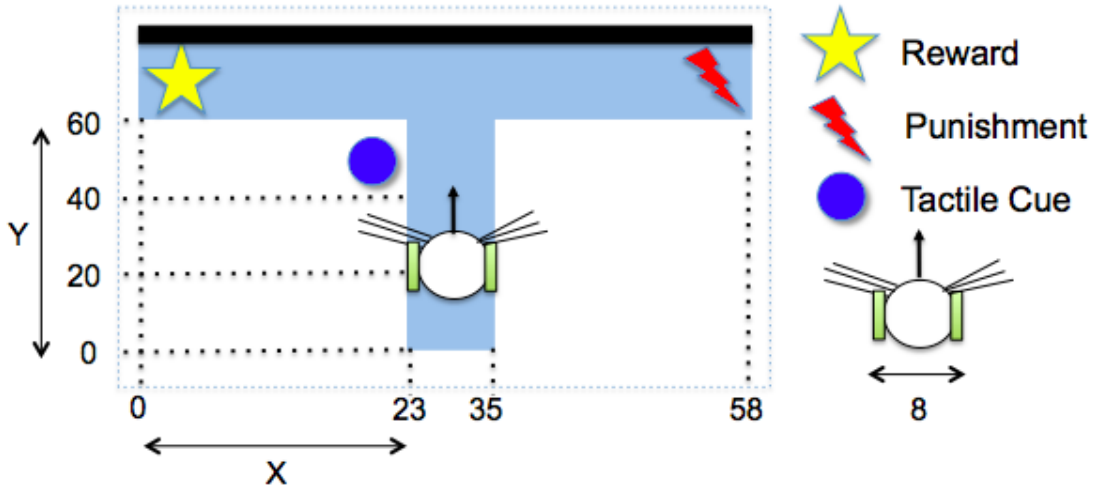

Figure 2-1: **T-maze environment used for the experiment.** At the beginning of each trial, the robot was placed on the central arm of the maze. The initial position of the center of the body was originally set to $(29, 20)$. The circle at the choice point represents the tactile cue, the star at one end of the maze indicates reward, and the lightning at the other end of the maze stands for punishment. All the sizes and distances are in centimeters, which is scaled based on actual e-puck robots.

The environment for this study was a T-maze with one central arm and two side ones, which is shown in Figure 2-2 along with its length and detailed arrangement. All the sizes and distances in the simulated environment are in centimeters, which is scaled based on actual e-puck robots [71]. A reward was located at the end of one arm, and a punishment was placed

at the end of the opposite one. The task of the robot was to reach the reward by choosing the right arm to follow. The right direction was indicated by a tactile cue at the intersection. The robot had to learn the correlation between the cue and the reward in order to complete the task successfully. Additionally, the robot also had to learn to move around within its environment. The robot was equipped with the following sensors and motors in the simulation environment:

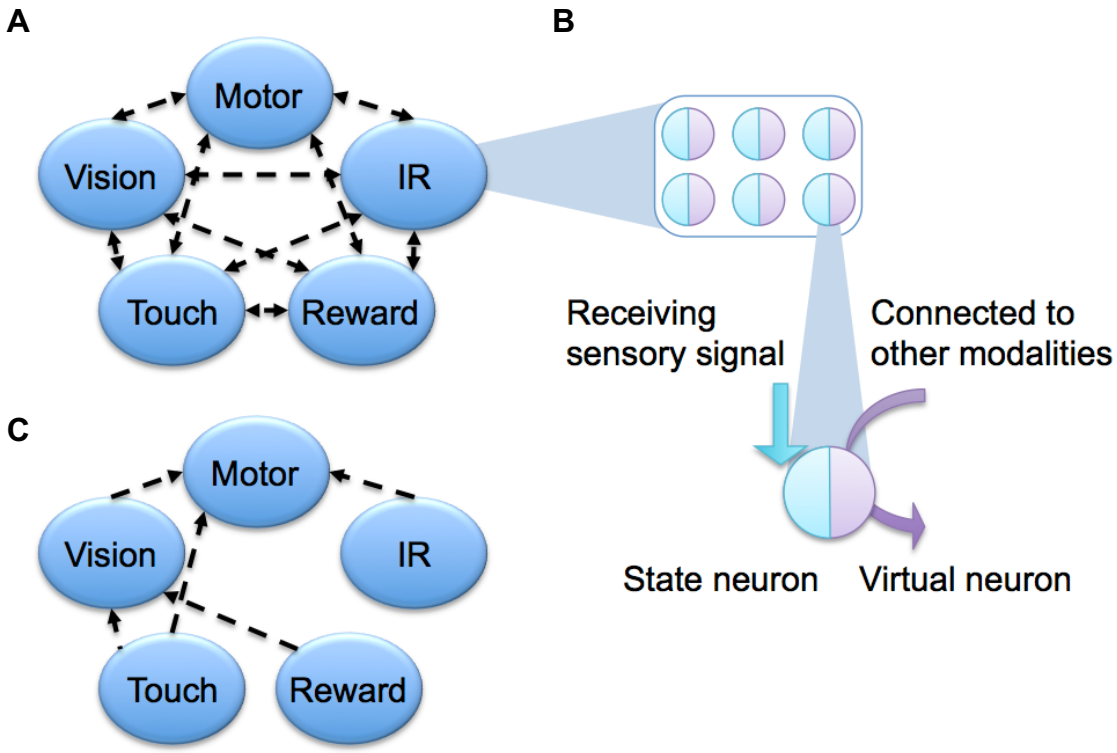

Figure 2-2: **Structure of the neural network. A:** Five sub-systems for each sensor or motor modules make up the whole cognitive system of the robots. These modules are fully connected with each other. **B:** One module is composed of several neurons. Each neuron has two components; one is a state unit, while the other is a virtual unit. **C:** Modules are minimally connected.

1) Tactile sensors: Tactile stimulation came from the 32 whiskers attached to the left and right sides of the robot. The length of the whiskers was 20.0 cm (Figure 2-2). The signals from those whiskers were given as the binary numbers 1 (triggered) or 0 (at rest). The whisker sensors only detected the tactile cue at the corner of the T-maze when the robot was close enough.

2) Vision sensors: Visual stimulation reflected the activity of the omnidirectional camera, which returned grayscale values standardized from 0 to 1. This camera was composed of 20 pixels aligned horizontally. Everything in the T-maze was made white or transparent except for the black backside wall (Figure 2-2). Therefore, only this wall provided a signal of 1 when entering the field of view of the camera. By sensing this black wall, the robot could estimate its location and direction within the arena.

3) Infrared (IR) proximity sensors: Six IR proximity sensors were uniformly attached to the front half of the robot's body. These sensors detected the distance from the robot to the walls of the T-maze.

4) Reward sensitivity: The reward sensitivity was usually set to 0. It was raised to 1 to signal a reward and lowered to -1 to indicate punishment.

5) Motors: The forward velocity of the robot was set to a constant positive value (denoted by v_1). The turning degree was determined by the output of the neural network v_2, which was standardized between 0 and 1. (v_1, v_2) were used to set the left and right wheel velocities as follows:

$$\begin{pmatrix} v_l \\ v_r \end{pmatrix} = \begin{pmatrix} v_1 + Cv_2 \\ v_1 - Cv_2 \end{pmatrix}, \tag{2.1}$$

where C was a constant for converting the standardized value to the actual motor speed. If $v_2 > 0$, then $v_l < v_r$, which made the robot turn left, and $v_2 < 0$ produced a right turn.

All the experiments described in this paper took place in two phases: a training phase and a testing phase. During the training phase, the robot explored the maze with the cue and reward deactivated and with the motor speed fixed to the two variables ($v_2 = (1, 0)$ or $(0, 1)$, which induce left and right turn respectively). When it touched a wall, the robot was returned to its initial position and the motor speed was switched to the other value. This process was repeated 20 times. The purpose of this phase was to give the robot a chance to learn the correlations between the IR, vision, and motor modules. After the training phase, the robot started the testing phase, which consisted of the robot seeking the reward and learning the correlation between the

positions of the cue and of the reward, as described above (for more details about the learning processes, see supporting information; Text S1 and Figure S1–6). The testing phase had 100 trials, where one trial was finished if the robot reached the reward or punishment or was timed out (after 4,000 time steps). The training phase facilitated the testing phase, as the robot had to learn the basic sensor and motor correlations necessary for moving within the maze. The results presented later on only concern the testing phase.

2.2.1 Neural Network

In this section, we describe how the sensory or motor modules temporally updated their neural states. It is important to note that the neural network in this work is only intended as a behavioural model, rather than as a model of biological neurons. The goal of this simplified model is to focus on a dynamical structure emerged only from the interaction between body and environment, which we expect will reveal an essential mechanism that causes VTE.

Just for simplification, the description here is different from that of the original model [5], but it is still mathematically equivalent to the original version and satisfactory in terms of replication of the model. Each module had a specific number of neurons (tactile 32, vision 20, IR 6, reward 1, motor 1), which were connected to neurons in other modules. The respective neurons were composed of two units: state units and virtual units. The state units were set proportional to sensory signals obtained from each sensors (and motors), and used to tune the Hebbian synaptic weights. The virtual units provided a parallel pathway for sending signals to other modules, which were activated by internal signals sent from other modules (illustrated in Figure 2-1B). The reason for having the two types of unit was to separate the dynamics of Hebbian learning and from the internal neural dynamics. The separation into two types of unit was necessary in order to avoid an unwanted positive feedback in which simultaneous firing leads to a strengthening of the connection, which results in a greater likelihood of simultaneous firing in the future. Although the network structure that we used here was specific to the Bovet-Pfeifer model, we think the obtained results do not depend on the detailed architecture of the model. The separation of Hebbian learning from the internal dynamics is the most important in order to reproduce our result. Both types of units were activated as follows:

1) State units, to compute the Hebbian synaptic weights: State units $\boldsymbol{x}^m(t)$ were just set proportional to the sensory signals. For instance, tactile stimuli from the 32 whisker sensors gave sensory values (0 or 1) to the corresponding 32 nodes of $\boldsymbol{x}^m(t)$. The proportional constants for the light, touch and reward modules were determined by the genetic algorithm explained below. That of the IR was fixed to 1.0. Depending on the state units, the weight matrix from module m to n ($= W^{mn}(t)$) was updated using a modified version of the Hebbian learning rule:

$$\begin{aligned} \Delta W^{mn}(t) &= l \cdot (\boldsymbol{x}^n(t)\Delta\boldsymbol{x}^m(t)^T - \alpha|\Delta\boldsymbol{x}^m(t)|W^{mn}(t)), \\ W^{mn}(t+1) &= W^{mn}(t) + \Delta W^{mn}(t), \end{aligned} \tag{2.2}$$

where l was the learning rate, α was the forgetting rate, and $\Delta\boldsymbol{x}^m(t)$ was the difference between the current and delayed state units ($= \boldsymbol{x}^m(t) - \boldsymbol{x}^m(t-\tau)$). This equation means that the weights from module m to n were strengthened when signals of module n and the signal change of module m were both high. In other words, the weight matrix encoded associations between sensory inputs, motors and reward signals. The reason why the signal change was taken into account is because it often has meaningful information rather than the state itself, as is explained by [5]. For instance, the motor state could be correlated with optical flow rather than a stable visual image.

2) Virtual units, to send signals from one module to another: Depending on the Hebbian synaptic weights computed above, the virtual units $\tilde{\boldsymbol{x}}^m(t)$ were activated by signals sent from the other modules by the following equation.

$$\tilde{\boldsymbol{x}}^m(t+1) = f(\Sigma_{n\neq m} W^{mn}(t) \cdot \Delta\tilde{\boldsymbol{x}}^n(t)), \tag{2.3}$$

where $f(x)$ was a sigmoid function ($= 1.0/(1.0+e^{-ax})$) and $\Delta\tilde{\boldsymbol{x}}^n(t) = \tilde{\boldsymbol{x}}^n(t) - \boldsymbol{x}^n(t)$, i.e., difference between the virtual units and the state units of the nth module. The virtual units did not have any effect on the learning rule, but they sent signals to the other modules through the learned synapses. As an exception, the virtual unit of the reward module ($\tilde{\boldsymbol{x}}^{reward}(t)$) was not subjected to the update by Eq. 2.3 but constantly set to 1. The

other virtual units than the reward module and all of the synaptic weights were set to 0 at the beginning of both the training and the testing phase. The virtual unit of the motor module was equivalent to v_2 at the equation 2.1 to update the motor velocity.

The reason that the update function was based on $\Delta\tilde{\boldsymbol{x}}^n(t)$ (i.e., the difference between the state and virtual unit) is discussed in the supporting information (Text S1 and Figure S1–S4), and here we just give a summary. The virtual unit can often be interpreted as an ideal state [4], in the sense that the the robot's motion tends to make an actual state closer to the virtual (or ideal) one. In the reward module, the virtual unit was constantly set to 1.0 as explained above, and the robot spontaneously chose actions to make the state unit closer to it, i.e., receiving the reward (Text S1 and Figure S3). For the IR module, the virtual unit give rise to wall avoidance behavior as described in Text S1 and Figure S1.

All aspects of the model were updated every N time step, with the exception of the reward states and the robot's velocity, which were updated every time step. N was a parameter between 10 and 30, determined by the GA. The following are the steps that led to the generation of the outputs:

1. Sensory information was transferred to the state units.
2. The Hebbian learning rule was applied on the weight matrix $W^{mn}(t)$ depending on the state units (equation 2.2).
3. The activities of the virtual units were updated (equation 2.3).
4. Finally, the virtual units of the motor module $\tilde{\mathbf{x}}^{motor}(t)$ determined v_2 to calculate the motor output (equation 2.1).

2.2.2 Set-Up of the Genetic Algorithm (GA)

Bovet's robot relied on the following parameters: learning rates l and forgetting rates α for each directed pair of sensor or motor modules (${}_5P_2 \times 2 = 40$; equation 2.2), update frequency of the neural network N, τ for the delayed states, the constants a for the sigmoid function, C at equation 2.1 (for each training / test phase), the forward velocity v_1 (equation 2.1), and the

proportional constants for the light, touch, and reward module. The authors did not mention how one should select these parameters. However, we found that slight differences in their value strongly influenced the performance of the robot. To tune these parameters and optimize the performance of the controller, we employed a standard genetic algorithm [46]. we used a population of 100 individuals to optimize the 49 parameters and used a tournament selection with a single point cross-over operation with a probability of 70%, and 1% mutation rate. we also applied elitism by simply copying the 5 best individuals to the next generation without applying a mutation. The fitness function $F(t)$ of an individual evaluated at the generation t was calculated as:

$$F(t) = \begin{cases} +5 \text{ points, if it reaches the reward.} \\ +0.25 \text{ points, if it reaches the punishment.} \\ +0 \text{ points, if it gets timeout.} \end{cases} \tag{2.4}$$

The fitness rewards were determined experimentally. The trials were repeated 100 times from one fixed initial position, which gave a maximum fitness value of 500.

2.2.3 Plausible Mechanism to Generate the Reward Seeking Behavior

As was described above, we found several possible parameter sets by GA to maximize the fitness value. Although those different parameter sets might result in various strategies, here we briefly explain the most plausible mechanism to generate the reward seeking behavior of the robot as explained in [5, 6]. For more details, see supporting information (Text S1 and Figure S1–S4). It seems plausible that the reward seeking behavior was generated from the following two neuronal paths learned by Hebbian rule:

Reward – vision – motor correlation: This path leads the robot to go to the previous reward position. This can support the reward seeking behavior at the earlier stage of the T-maze task, i.e., before acquiring the meaning of the tactile cue.

Tactile – motor correlation: This path lets the robot follow the tactile cue to make a correct decision. At the beginning of the task, nothing about this path is not learned, but will gradually be acquired by moving around the environment.

The position of the reward changed every five trials of the T-maze task, so that the trials after switching the reward position could give a conflict because of the two neuronal paths described above. Namely, the first neuronal path would lead the robot to go to the past and wrong reward position. The second path would tell the robot the correct reward position suggested by the tactile cue. Therefore, before maturation of the synapses, this experimental setup could cause a conflict-like behavior, or VTE. However, depending on the parameters, the robot could still solve the maze by different learning procedures.

2.2.4 Neural Network - Minimally Connected Model

As the task presented in this paper is simple conditional learning, it can be performed using a minimal network topology. In the neural network described above, the sensor and motor modules were fully connected, which could be redundant to solve the task. On the other hand, the minimal network we propose only had connections with a specific role to solve the task, which was suggested by Bovet himself in [5]. In Figure 2-1C, the minimal network is described, which has only the four neuronal paths; “touch – vision”, “IR – motor”, “touch – motor” and “reward – vision – motor”, the last two of which were described above. we compared the behavioral difference between the evolved but redundant network and the minimally connected network to show the relevance of redundancy with VTE.

2.3 Results

As we have described in the previous section, we examined two types of neural networks in the robot – a fully connected one and a minimally designed one. For the fully connected network, we ran the GA 95 times. Of these, 22 runs produced a network that attained a maximum fitness value (100% success). For the minimally connected network, 2 out of 5 runs of GA attained the maximum fitness. we first counted the number of VTEs associated with these evolved robots.

In his experiment [108], Tolman counted the number of VTEs through the oscillations of the rat’s head direction; i.e., one VTE was counted if the rat looked at one door and then it looked at the other door. we followed this method for counting the number of VTEs in the robot. In this experiment, since the robot did not possess an independent head from its body, we counted the

left and right oscillations of the whole body as VTEs. More precisely, a VTE was granted when the motor output v_2 from the equation 2.1 changed its sign. In order to filter noisy fluctuation around the turning degree of 0, we counted a movement as a VTE when the sign change of v_2 was larger than the threshold range $[-T, T]$, which was set to $T = 0.3$. As we discuss later, the threshold $T = 0.3$ was varied to see the role of VTE.

The evolved robots showed different patterns of VTE that can be classified into three groups. The classification was done by hand, and five out of the 22 evolved robots could not be classified as they showed exceptional form of VTE. Therefore, we had 16 robots to be classified. Some examples of VTE dynamics exhibited by individual robots are shown in Figure 2-3, and those of all the 16 robots are shown in Figure 2-13 (black lines), which will be shown below in this paper. The three groups are as follows:

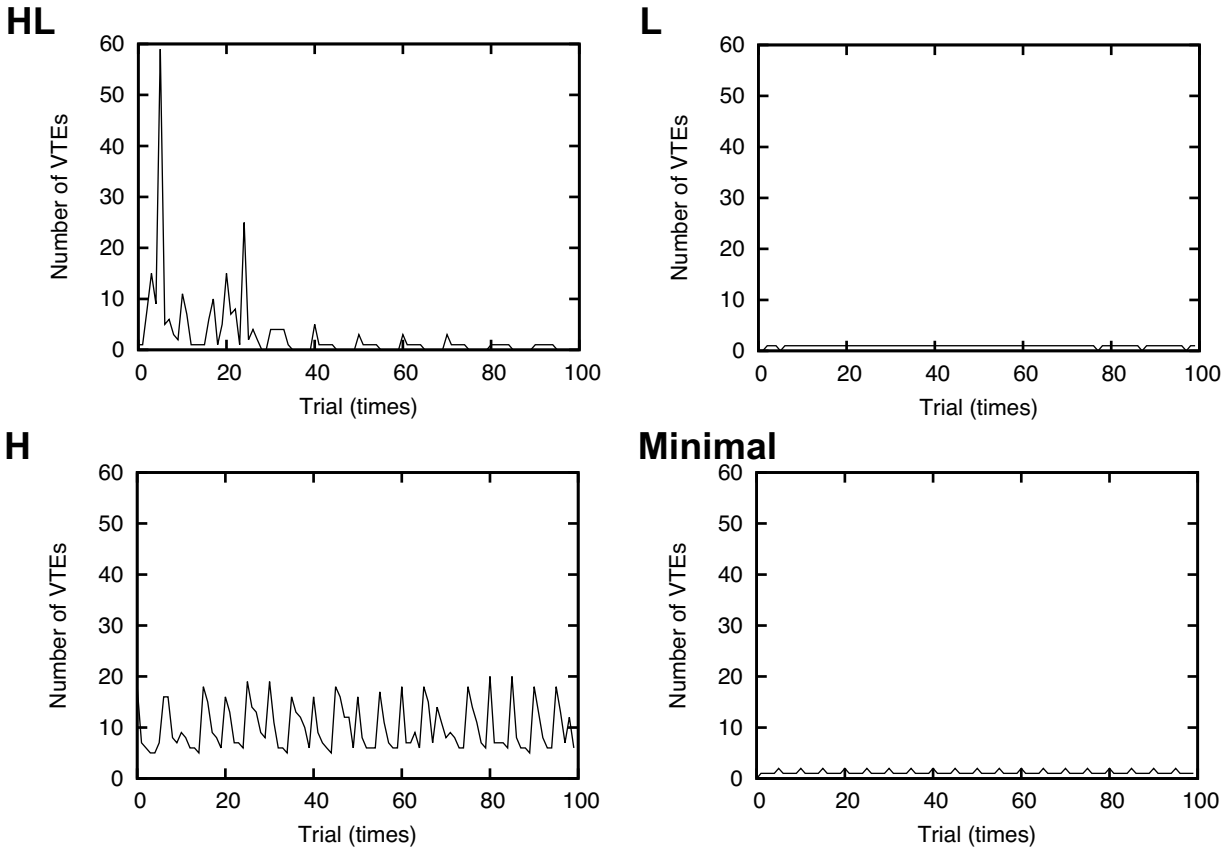

Figure 2-3: **Examples of the number of VTEs shown by individual robots during the 100 trials.** The X axis indicates the trial numbers, while the Y axis show the number of VTEs. **HL:** High to low VTE. **L:** Low VTE. **H:** High VTE. **Minimal:** Minimal model.

A high to low number of VTEs (denoted by "HL") The number of VTEs increased at the early stages and then gradually decreased after a certain point (Figure 2-3HL and 2-13HL). This observation was similar to the experiments with rats [108, 111]. we had four robots out of the 16 that showed this type of VTE. For the robot shown in Figure 2-3HL, the number of VTEs had its peak at 6th trial, when the position of the reward changed. As for the other three evolved individuals in Figure 2-13, the second one also showed the VTE peak at the 6th trial. The third and the forth one took the peak at 7th. Therefore, VTE had its peak just after the reward switching event. This peak disappeared later on after the robot had more time to learn the task.

Low VTE (denoted by "L") The number of VTEs was kept constantly low throughout the 100 trials (Figure 2-3L and 2-13L). Six robots exhibited this type of VTE.

High VTE (denoted by "H") The number of VTEs was kept high (but temporally fluctuating) throughout the 100 trials (Figure 2-3H and 2-13H). Five robots had this type of VTE. Two out of the five robots had local maxima in the number of VTEs when the reward position switched, while the other three did not.

The two robots with minimal connectivity (denoted by"Minimal") were all classified as having low VTE (Figure 2-3 Minimal). This suggests that the redundant connectivity between the sensor and motor modules was necessary. In the next part, we investigated the mechanism of VTE to show that VTE was generated from a chaotic neural activity.

As is shown in Figure 2-13, there is a continuum between HL and H robots, while the robots of L (and Minimal) seem distinct from those two VTE-showing groups. Despite this difference in the number of VTEs, each robot showed a perfect success rate (of 100%), which implies that VTE is not directly related to the degree of performance in the T-maze task. This seems inconsistent with what is observed in animal experiments [47, 115], where VTE is efficient for learning performance. Still, bellow we suggest that VTE is efficient for robust learning under environmental perturbation. In the remainder of this paper, we describe a possible mechanism for generating VTE and then evaluate the function of VTE in terms of robustness and dynamic stability. Finally, we discuss the effect of VTE on Hebbian learning.

2.3.1 Mechanism behind VTE

Figure 2-4 shows an example of the temporal dynamics of the averaged synaptic weights in the cases of HL and L. This entire time series consisted of 100 successive trials, with each trial being comprised of roughly 15 to 150 neural time steps (i.e., measured by the number of updates of the neural states) of starting from an initial position and ending at either end of the T-maze (or running out of the assigned time duration). Here, we investigated the neural mechanism that generates VTE.

The trajectories of the robot during the 100 trials are shown in Figure 2-5. It can be seen that the HL robot showed unstable trajectories (Figure 2-5HL), while the L robot (Figure 2-5L) showed stable ones. This observation suggests that the orbit that the HL robot took was destabilized, while the orbit of L was stabilized.

Therefore, we hypothesized that VTE is generated from chaotic activity of the neurons. To quantify the degree of chaos, we computed the maximum Lyapunov exponents (MLE), or an index of chaos, by quantifying the instability of the orbits, which is defined by;

$$d(t) = d_0 e^{\lambda_1 t}, \tag{2.5}$$

where d_0 is the distance between the two neighboring points, while $d(t)$ is the average divergence between the two after t time steps. If the $d(t)$ exponentially increases with time t, then the λ_1 converges to a positive value, where the system is regarded as chaotic. Rosenstein et al. provided an algorithm for calculating the MLE from an experimental time series [89]. The MLE was computed by the four steps:

1. Reconstruct k dimensional vectors $\mathbf{x}_i = \{x_i, x_{i+\tau}, ..., x_{i+(k-1)\tau}\}$ from the given time series $\{x_1, x_2, ..., x_N\}$, where τ represents the time delay.

2. Choose the closest pairs of the reconstructed vectors and compute how the distance between the two develops over time. By using the fast Fourier transform (FFT), the mean period T is computed, where closest pairs are selected from vectors that are not in the same period. $d_j(i)$ denotes the distance evolved after ith time steps between the jth pair of the closest pairs. Distance between two vectors ($\mathbf{x_i}$, $\mathbf{x_j}$) is calculated by Euclidean norm

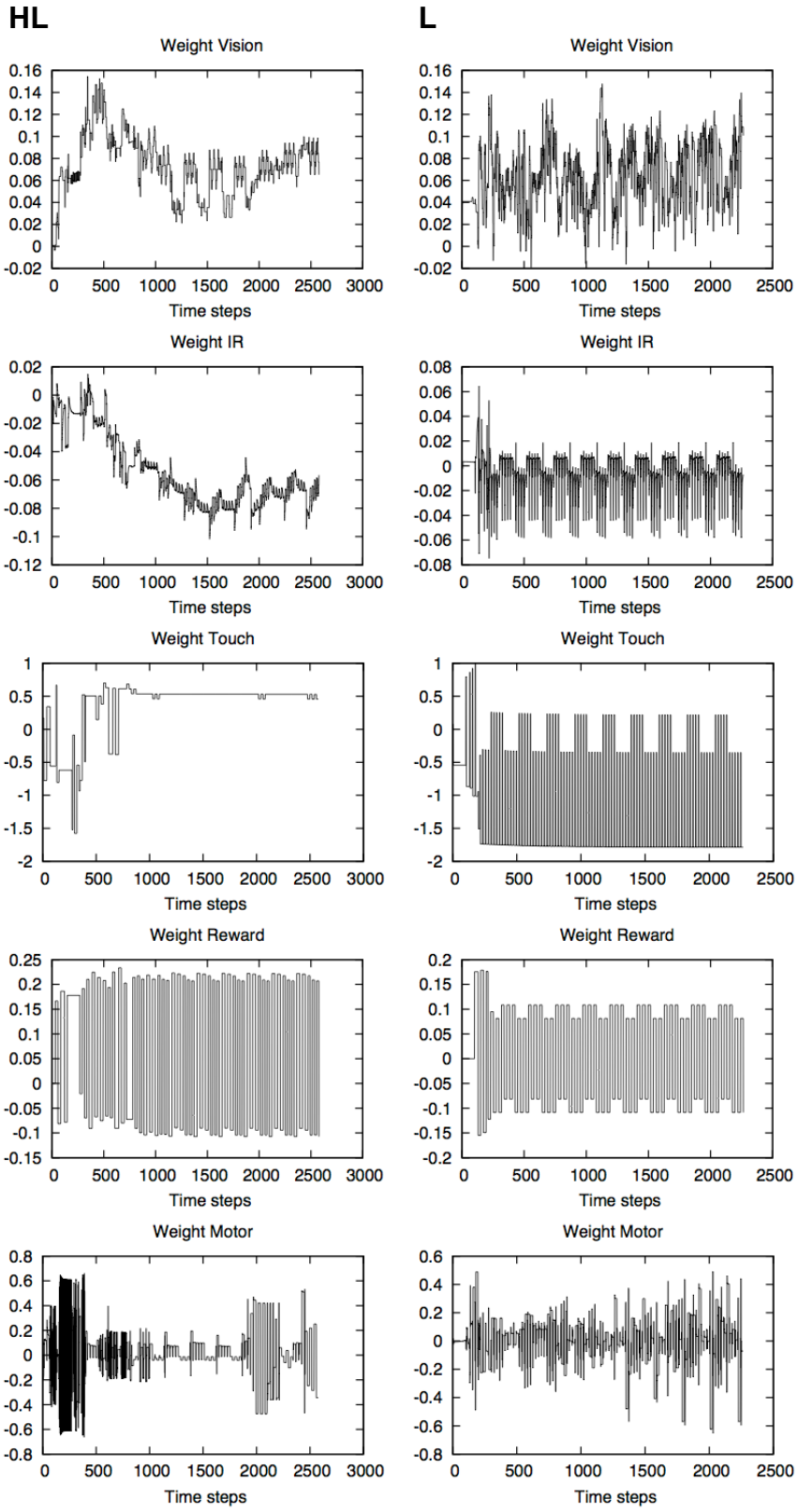

Figure 2-4: **Examples of synaptic weight dynamics of the HL and L during 100 trials.** The X axis denotes time steps, while the Y axis shows the strength of the synaptic weights. **HL:** High to low VTE. **L:** Low VTE.

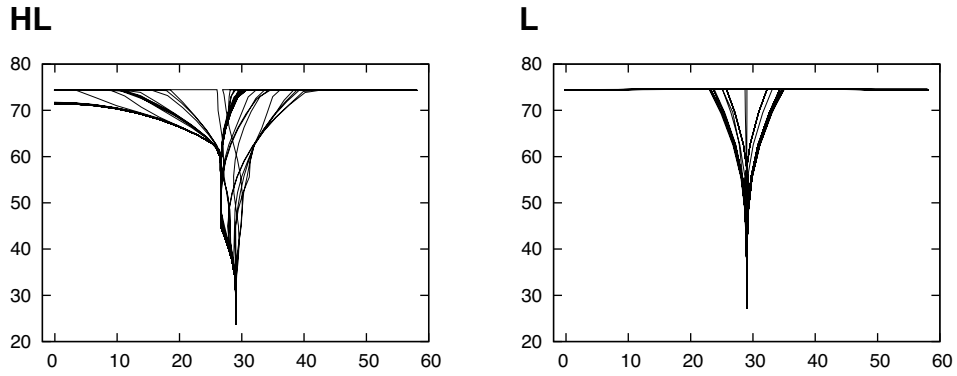

Figure 2-5: **Trajectory of the robot.** Each trajectory of 100 trials is superposed. **HL:** High to low VTE. **L:** Low VTE.

$\|\mathbf{x_i} - \mathbf{x_j}\|$.

3. The divergence rate of $d_j(i)$ is approximated as:

$$d_j(i) \approx C_j e^{\lambda_1 \cdot i}, \tag{2.6}$$

where λ_1 denotes the MLE. The equation above can be rewritten by:

$$\ln d_j(i) \approx \lambda_1 \cdot i + \ln C_j. \tag{2.7}$$

The logarithm of the distance d is proportional to λ_1, where we can estimate the value of the MLE. Therefore, λ_1 is estimated from a least-squared fit to the average line ($y = \langle \ln d_j(i) \rangle$), which is obtained from sampling the $N - (k-1)\tau$ pairs.

In the current experiment, we adopted the slope of the fitted line as the MLE only when the R-squared value (i.e., the goodness of fitting) was larger than 0.8, and otherwise the MLE was set to 0.0. When the estimated value was negative, then the value was also set to 0.0. In general, negative Lyapunov exponents indicate the convergence speed of the two orbits, which is insignificant after reaching a stable state.

The synaptic weights were not reset through the 100 trials, so the robot started each trial with different initial weights (since it inherited the synaptic weights from the previous trial). We thus fixed the synaptic weights to a initial weight set for each trial, and let the robot move in the

T-maze without learning, where we obtained time series of neural activity to compute MLE. This setup was to understand dynamical structure determined by the weight sets for respective trials. We acquired roughly 15–150 data points of the neural activity from the run and obtained the time series of the averaged neural activity for respective sensor or motor module. We used the virtual unit $\tilde{\boldsymbol{x}}^m(t)$ to create the time series for the motor module and the differences between the virtual and state units $\Delta\tilde{\boldsymbol{x}}^m(t)$ for the other sensory modules. This is because the motor value v_2 (equation 2.1) was equivalent to the virtual unit of the motor $\tilde{\boldsymbol{x}}^{motor}(t)$, so that it was more important than the differences $\Delta\tilde{\boldsymbol{x}}^{motor}(t)$ for determining the robot's behavior. For the other sensory modules, the differences $\Delta\tilde{\boldsymbol{x}}^m(t)$ propagated to the motor module to determine v_2, so that they were more effective on its behavior.

Figure 2-6 shows several examples of time series that were used for MLE computation. These time series were obtained from the $32th$ trial as examples, for all of the 16 evolved robots. The X axis shows time steps, while the Y axis indicates the activity level of neurons. Colors represents sensory or motor modules, i.e., vision: gray, IR: purple, touch: green, reward: red, motor: brown. This figure indicates that the neural activity appeared different depending on the type of VTE. The HL and H robots showed oscillating and unstable neural activities, while the L and Minimal robots exhibited rather stable activities.

Figure 2-7 shows the averaged MLE over the 100 trials, where the error bar indicates standard deviation. We set $k = 4$, and $\tau = 1$. The MLE is calculated for the respective five modules of each of the 16 evolved robots, which are denoted by HL1, HL2, ..., Minimum2. We can see that HL and H robots showed positive MLEs more often than the L and the Minimal robots. This suggests that 1) Hebbian learning leads to both chaotic and non-chaotic neural activity even in the same environment and 2) chaotic activity is positively correlated with the presence of VTE, while non-chaotic activity is not.

In this part, we mention validity of the MLE estimation. We used time series that consists of around 15–150 data points, which is quite small compared with a conventional way of MLE estimation. For instance, Rosenstein et al. estimated the MLE of Logistic map, Lorentz attractor, or Rössler attractor, by using 500–5000 data points [89]. Enough data points are needed to see if the state converges into an chaotic attractor, because chaos is defined as an attractor that has positive MLE. In this respect, our data might not have enough points to prove the convergence.

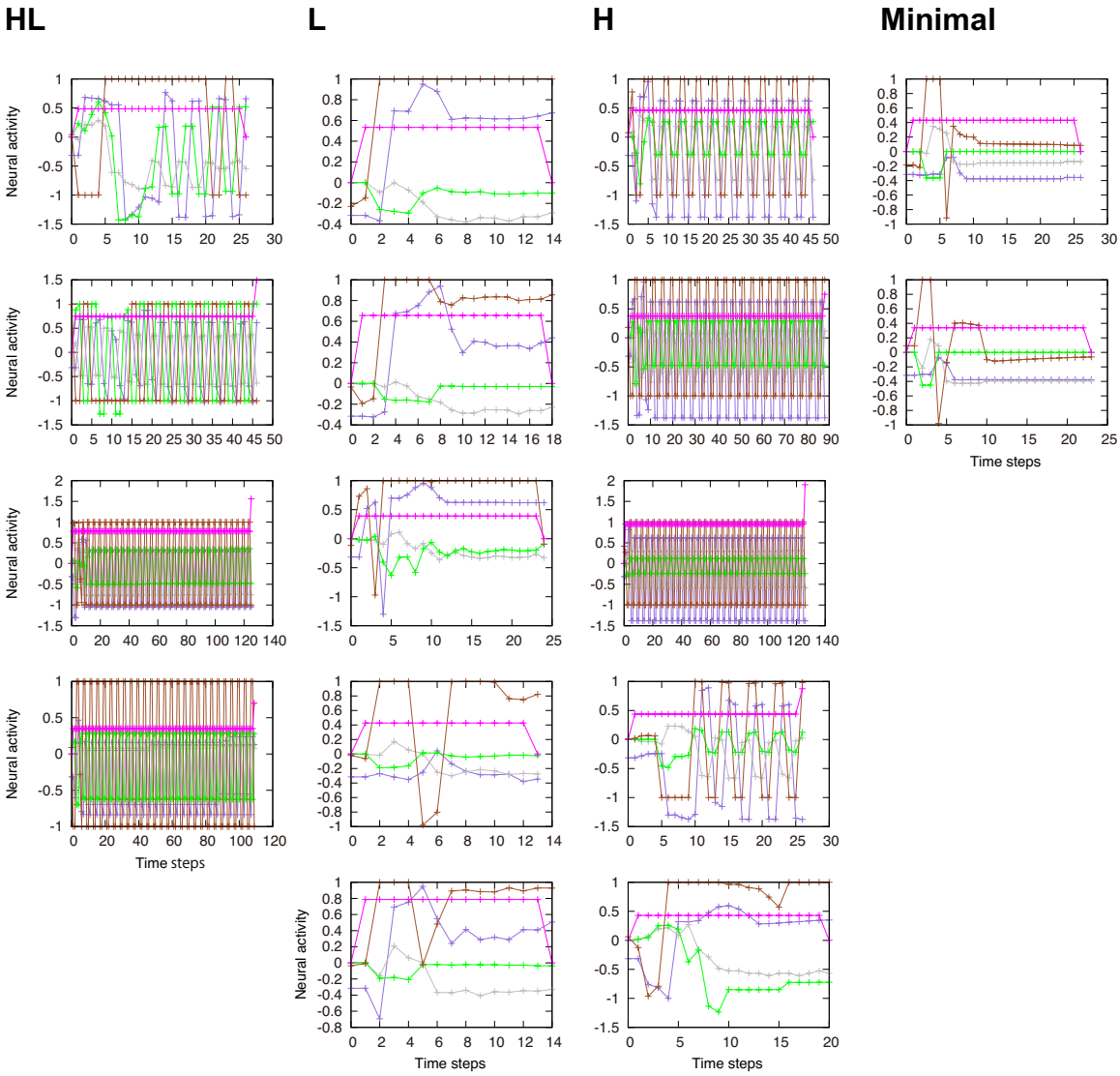

Figure 2-6: **Examples of time series used for the MLE computation.** The time series presented here is obtained from the $32th$ trial as examples, for all of the evolved 16 robots. The time series is the averaged neural activity within a same module. The virtual unit $\tilde{\boldsymbol{x}}^m(t)$ for the motor module and the differences between the virtual and state units $\Delta\tilde{\boldsymbol{x}}^m(t)$ for the other sensory modules are presented. The colors of the lines represent sensory or motor modules, i.e., vision: gray, IR: purple, touch: green, reward: red, motor: brown.

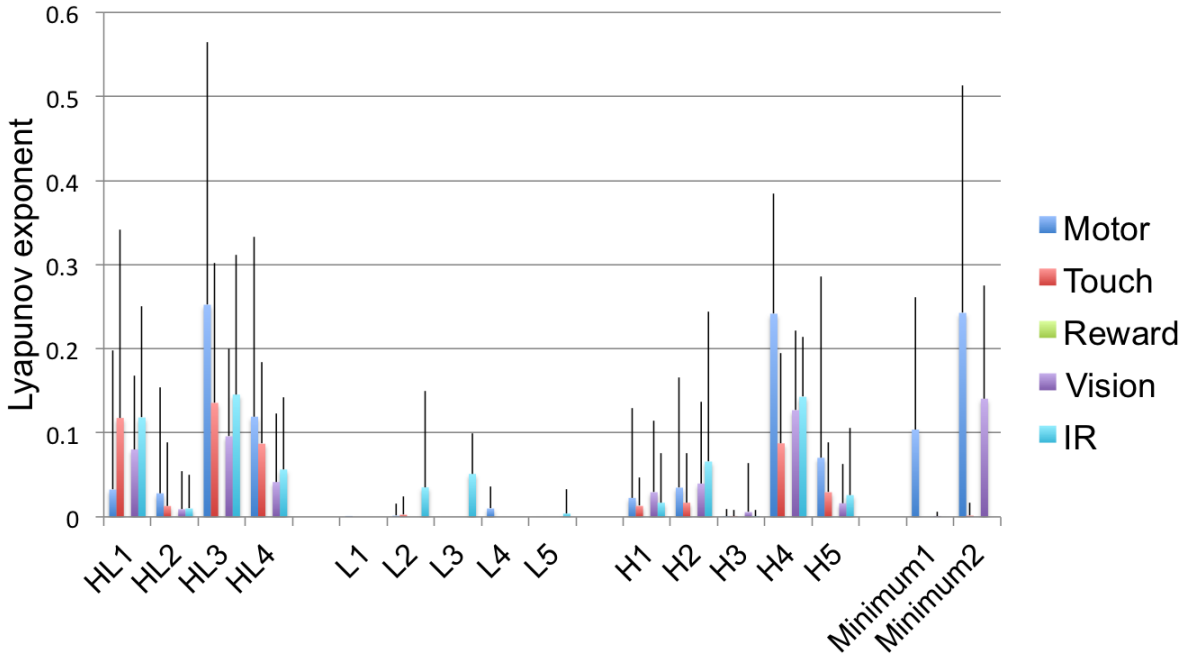

Figure 2-7: **Maximum Lyapunov exponents (MLE) averaged over the 100 trials.** The MLE is calculated for each five modules, and for each evolved 16 robots, which is denoted by HL1, HL2, ..., Mimimal2. The time series used for estimation the MLE is an averaged neural activity for the respective modules. The error bar indicates standard deviation.

However, the MLE measured the divergence rate between neighboring pairs, so that it was still able to explain the complexity of the time series. Indeed, the MLE that we obtained appeared to explain well the time series as presented in Figure 2-6. Although our data were weak to prove chaos, we mention the system with positive MLE as being "chaotic" in the following parts of this chapter.

Note that the performance of each robot was very high (100% success) irrespective of the type of VTE. This might be seen as paradoxical because one might expect the instability to break the learned sensor and motor mappings that give successful behavior, resulting in low performance. Therefore, we hypothesized that the robot with VTE might actively use VTE to complete the task, not just as unstable head oscillations. This is evaluated in the next paragraph by calculating the robot's performance under perturbation of VTE.

The next step of our investigation was to determine whether VTE was a mere epiphenomenon or if it had a specific role in completing the task. We took a HL robot (HL1), which is the same one as that in Figure 2-3HL, and artificially prevented the presence of VTE at the motor level by resetting the angular velocity to zero every time VTE was detected, forcing the robot to maintain its current direction. Our hypothesis was that if VTE is a mere epiphenomenon and not necessary for achieving the maze task, preventing it will not alter performance. On the other hand, if VTE is necessary or at least helpful for the task, the prevention of VTE will decrease performance. In order to test the hypothesis, we prevented VTE with a threshold T. If the motor output v_2 (equation 2.1) changed sign and the change was larger than the threshold range $[-T, T]$, one VTE was counted. We varied the VTE threshold from 0 to 1, where threshold = 0.0 meant that any body rotation would be regarded as VTE and they were all prevented, while threshold = 1.0 meant that no body oscillations were prevented (the original case). The results of this analysis are shown in Figure 2-8. We found that introducing the threshold would always reduce performance. The results show the absence of VTE inhibited the correct acquisition of the task while maintaining an accurate control of the robot. Therefore, we considered that VTE is used actively to complete the T-maze task. This result is consistent with experiments of real rats and humans, where better performance is found in the presence of VTE [47, 115]. In the next section of this paper, the usefulness of VTE is linked to robustness of the behavior.

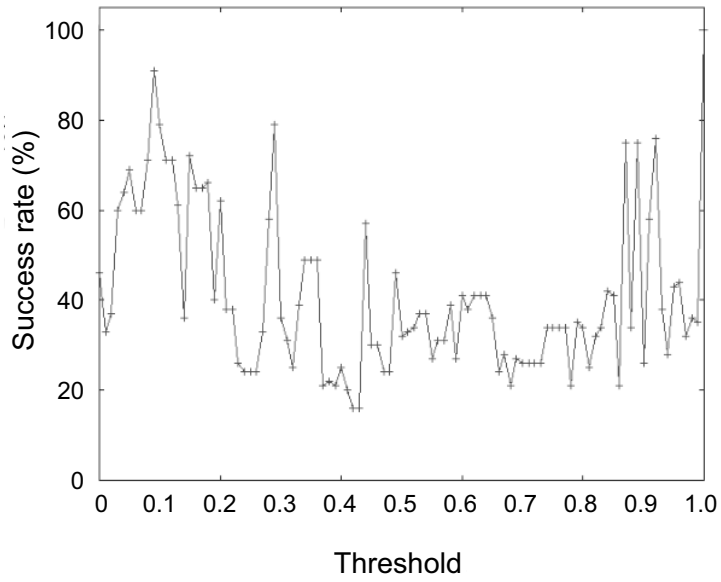

Figure 2-8: **Change in performance with blocked VTE.** The X axis represents the threshold. The Y axis stands for the success rates of the T-Maze task.

2.3.2 Robustness of VTE

To investigate the robustness of the robot's behavior for each types of VTE, we analyzed its performance under varying initial conditions. During evolution, the starting position of the center of the body was fixed to $(x, y) = (29, 20)$, as shown in Figure 2-2. This experiment explored whether perturbations to the starting position affected the performance by testing the robot from different starting positions inside the central arm of the T-maze. The robot repeated the task 100 times from different initial positions by changing (x, y) as $x = 27, 28, ..., 31$ and $y = 5, 6, ..., 60$ (280 initial positions in total). We calculated the success rates for each initial position.

Figure 2-9 shows the results of the HL and L, respectively. The X and Y axis indicate the coordinates of the central arm of the T-maze. Each pixel of the figure indicates the success rate when the robot started the 100 trials from this position. This figure shows that the performance was not constant for all the initial positions and allows a comparison of the variation in performance between the two models. Figure 2-9HL gives an example of HL, where the robot mainly obtained around 50% success rates, with several initial positions leading to success rates of 100% or below 20%. On the other hand, Figure 2-9L shows an example of L, where the robot

obtained mainly a success rates of 100% or 0% with several initial positions giving around a 50% success rate, suggesting a higher variance of the success rate. This tendency is summarized by Figure 2-10, which shows the average and the variance of the success rates for every evolved robot. Figure 2-10A indicates the average of the success rates, showing that the types of VTE do not have varying impacts on the average success rates. On the other hand, Figure 2-10B shows the variance of the success rates – the HL robots kept the variance under 400 (red line), while the other three groups had a variance above 550 ($p < 0.01$; student t-test). Therefore, despite the 4 groups had a similar average performance, the HL robots withstood changes in its initial position, while the other types were strongly affected by those changes. This result suggests that the presence of the HL type of VTE, which is similar to that of the rats [108, 111], is associated with a higher level of robustness. Animal experiments could be performed to test this hypothesis. Robustness acquired by VTE also explains why the prevention of VTE lowered the overall performance of the robots. Namely, preventing VTE will reduce robust control, and destabilize the behavior.

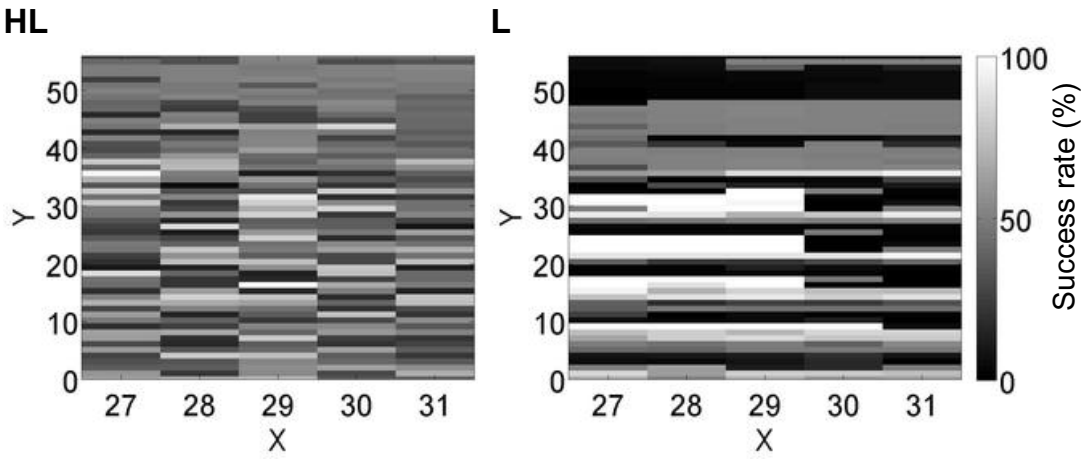

Figure 2-9: **Examples of the average success rates for each starting positions.** Each grid cell is filled with a colour representing the success rates when a robot starts the task from within the grid cell. **HL:** High to low VTE. **L:** Low VTE.

To further investigate the robustness of the evolved controllers against environmental change, we carried out the same experiments with different T-maze sizes. We varied the length of the width and height of the T-maze, and calculated the average and the variance of the success rates for every starting position. Figure 2-11 shows the variance of the success rate, where the HL

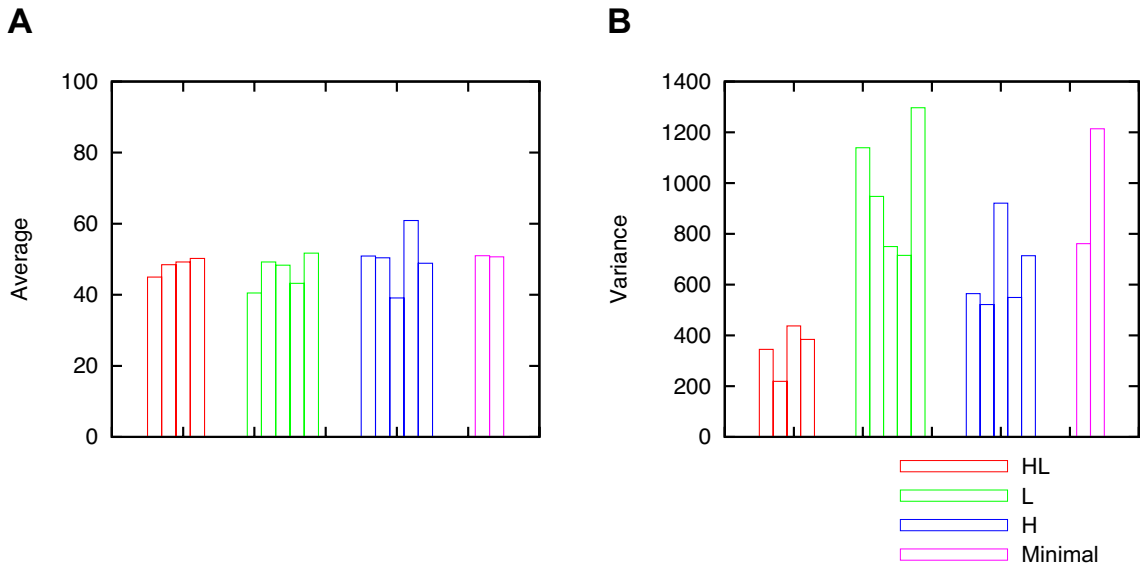

Figure 2-10: **Average and variance success rates over all the initial position** (N=280). Red, green, blue and pink bar indicate the results of HL, L, H and Minimal respectively. **A:** Average success rate. **B:** Variance of success rate.

robots did not change their performances with respect to the variance, while the robots in the other three groups were affected by a slight change in environmental size. This result confirms that the presence of VTE can be an indicator of the robustness of behavior.

2.3.3 Effective Use of Hebbian Learning

We have shown that VTE is used actively to achieve the T-maze task, rather than serving as mere oscillations of the head direction. This leads to the question of why the destabilized orbit by VTE guarantees the best performance of the robot. We hypothesized that the destabilization by VTE enables continuous learning of the environment, while too much stable dynamics found in non-VTE robots creates a loss of adaptability in learning. To investigate this hypothesis, we conducted the following two experiments:

1) First, we varied the position of the reward by moving it further away within the same branch of the T-maze. This means that the robot could not get the reward in the expected position. If the robot was continuously gathering information from the environment, then this set-up would create a contradiction between the expectation and the sensory information. On the other hand, if the robot was solving the maze just by reflex to the tactile cue, then it would

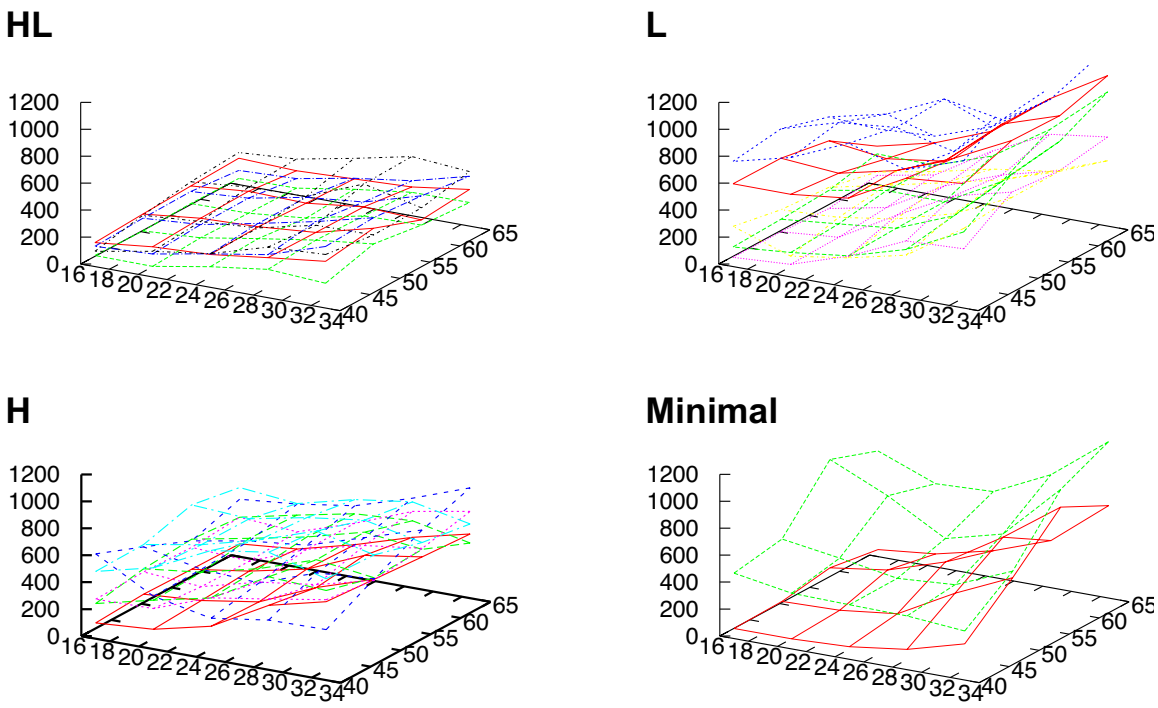

Figure 2-11: **Examples of the variance of the success rates with different environmental size.** The X and Y axis show the size of X and Y in the figure 2-2. The Z axis shows the variance. **HL:** High to low VTE. **L:** Low VTE. **H:** High VTE. **Minimal:** Minimal model.

be still successful, regardless of the reward position.

The results of this experiment are shown in Figure 2-12: all of the HL robots and some H robots reduced their success rates (Figure 2-12HL and H), while most of the L and Minimal robots maintained their high success rates (Figure 2-12L and Minimal). As these results suggest, the HL robots modified their behavior by continuously gathering information from their environment. On the other hand, the behavior of the non-VTE robots (L and Minimal) was only dependent on the cue command, and not on the reward position, so that they seemed just to obey the cue command, which is reflective rather than ongoing decision making.

2) Second, to evaluate the effect of learning, we compared the performance of the robots with and without learning. Originally, the synaptic weights were continuously updated through the 100 trials, so that the robot started each of the 100 trials with different initial weights. Therefore, we had 100 initial weight sets, one for each corresponding trial. In the no-learning condition, the robot started each 100 trials with an initial weight set corresponding to the trial, where the synaptic weights were fixed and not updated. In the learning condition, a robot

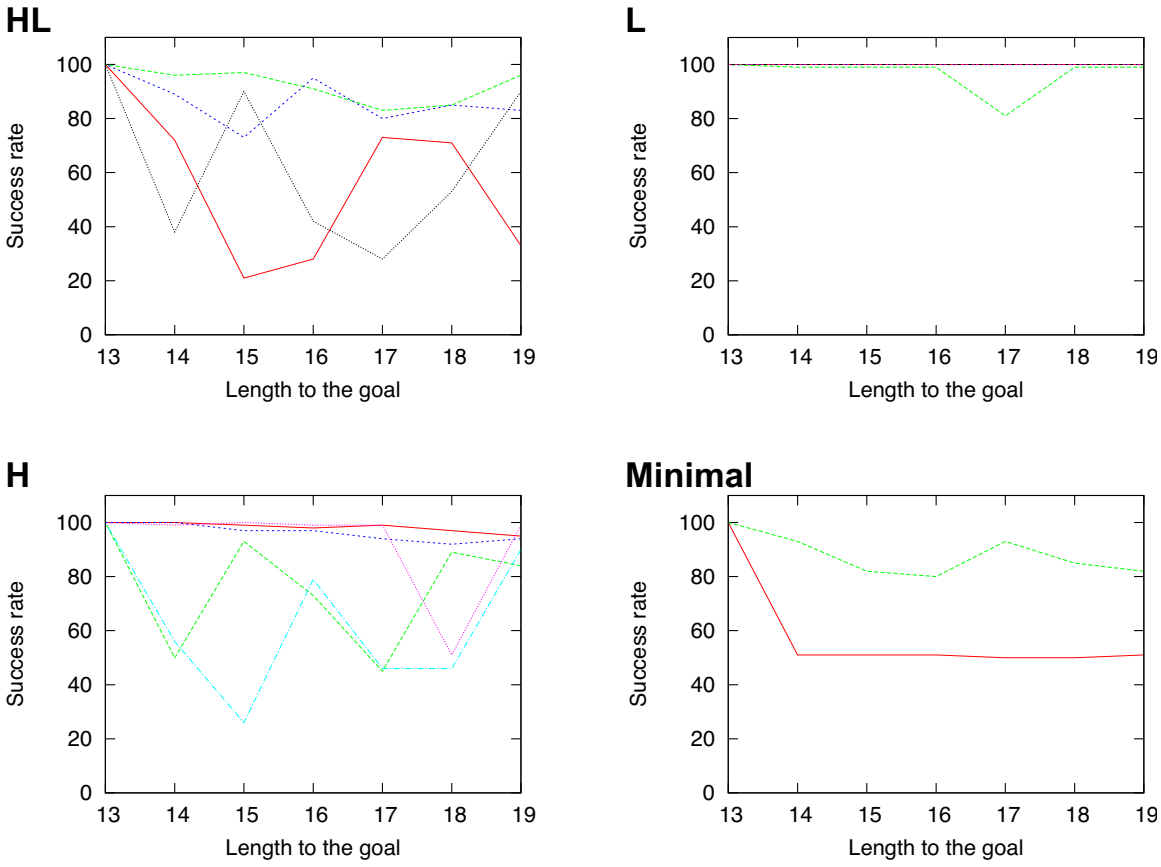

Figure 2-12: **Success rates with longer distance from the tactile cue to the reward.** The X axis shows the length to the goal, while the Y axis indicates the success rates. **HL:** High to low VTE. **L:** Low VTE. **H:** High VTE. **Minimal:** Minimal model.

updated its weights in the same way as the original experiment. For each trial, we computed the performance by perturbing the initial conditions, similar to what we did previously for computing the robustness; i.e., the initial position was changed for every grid inside the central arm of the T-maze.

Seventeen examples of these results are shown in Figure 2-13. The red line shows the success rates with learning, while the green line indicates success without learning. The black line indicates the number of VTEs observed in the original condition (i.e., with learning activated and from the original starting point; the same condition as Figure 2-3). For the HL and H robots, the performance was worsened or almost the same when the learning was deactivated. On the other hand, the L and Minimal robots improved performance without learning. This tendency suggests that the robots with VTE utilized learning to maintain their performance, while, for those without VTE, the learning process did not work properly.

We have therefore concluded that the destabilized orbit generated from VTE allows continuous learning by gathering information from the environment, which means that the robots are performing embodied cognition, i.e., ongoing generation of their behavior from interaction between the body and the environment. On the other hand, the robots without VTE are not embodied in their environment, which does not allow ongoing learning.

2.3.4 Multiple time scales observed in HL robots

To evaluate information flow in the neural network, we calculated transfer entropy (TE), which measures directed information flow from one node to another [2, 65, 95, 103]. For the definition of TE, see Chapter 3 Section 3.3.3.

To compute TE, we binarized neural activity into 0 and 1, where we applied it to the virtual unit $\tilde{\boldsymbol{x}}^m(t)$ for the motor module and the differences between the virtual units and state units $\Delta\tilde{\boldsymbol{x}}^m(t)$ for the other sensory modules. The threshold to banalize the raw data is currently set to 0.1. To evaluate information flow in different time scales, we compressed the binarized data with a different bin-length denoted by Δt. Namely, neural spikes (i.e., value 1) within the same bin were regarded as one spike. If $\Delta t = 1$, the obtained time series represents the original binary data, while a larger Δt represents the macroscopic behavior created by an ensemble of neurons.

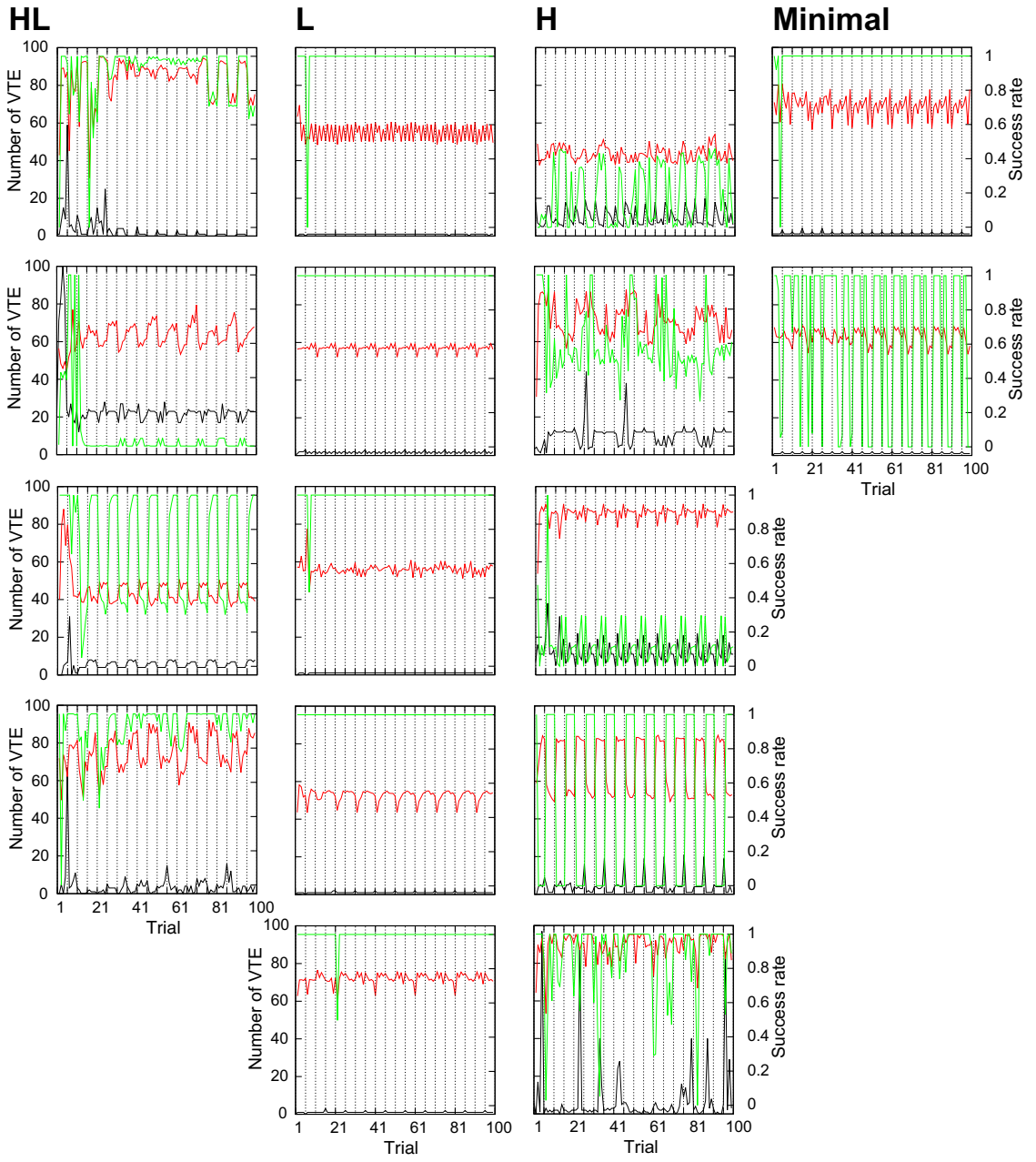

Figure 2-13: **The number of VTEs for the evolved 16 robots, and comparison of the performance between with and without learning.** Each figure represents the result from an individual robot. The scale of each axis is the same in every figure. Red line: Success rate when the learning is on. Green line: Success rate when the learning is off. Black line: Number of VTEs observed in the original condition (i.e., with learning activated and from the original starting point; the same condition as Figure 2-3). **HL:** High to low VTE. **L:** Low VTE. **H:** High VTE. **Minimal:** Minimal model.

Figure 2-14 shows examples of TE in the HL and L robots, where an directed edge is drawn from one neuron to another if the directed pair of neurons showed higher transfer entropy than a threshold (which is currently set to 0.04). The colors of the lines represent modalities of the pre-synaptic nodes: vision: gray, IR: purple, touch: green, reward: red, motor: brown.

These figures indicate that the HL robots show different topological structure depending on the time scales: For instance, edges from the reward and IR modules are observed with smaller Δt, while the motor module became dominant with larger Δt. On the other hand, the robot belonging to L group did not have noticeable peaks compared to those of HL and H.

Figure 2-15 shows ratio of the edges toward all the possible edges (vision:1200, IR: 360, touch: 1920, reward: 60, motor: 60). The subfigures in the figure represent the results respectively from the evolved 16 robots. The X axis represents time scale Δt, while the Y axis shows ratio of the edges over all the possible edges. The colors of the lines correspond to respective modules in the same way as in Figure 2-14 (i.e., vision: gray, IR: purple, touch: green, reward: red, motor: brown). From a simple observation of this figure, in the HL and H robot – i.e., VTE-showing robots, the ratio of the edges varied in accordance with Δt. On the other hand, the robot belonging to L and Minimal groups – i.e., non VTE robots – appeared not to show huge differences in the number of edges. This result indicates that regulation of proper time scales was more important in the VTE-showing robots.

2.4 Discussion

VTE is a behavior observed in experiments with rats that seems to suggest self-conflict [72, 73, 108]. It is mainly observed when rats are uncertain about making a decision; e.g., when they make a mistake or change their strategy [94]. The presence of VTE is regarded as an indicator of a deliberative decision making process; i.e., the process of searching, predicting and evaluating future outcomes [111]. Deliberative decision making is an opposite notion to an automated decision making process, such as habituation or reflex, and is computationally slower, while it can allow ongoing control to achieve flexible behavior. In fact, better performance is found when animals show VTE [47, 115].

In this paper, we have tried to show the underlying mechanism responsible for the obser-

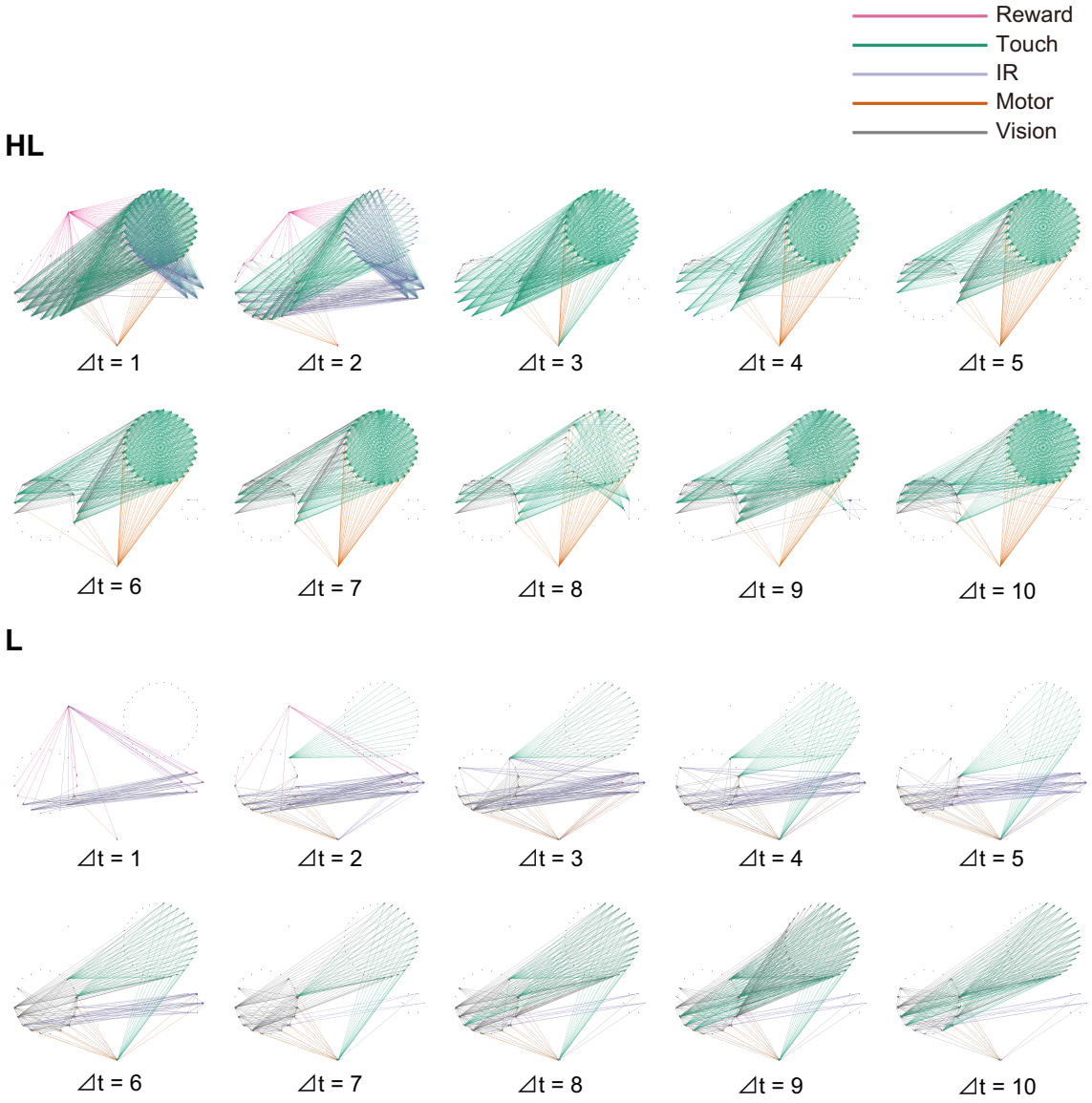

Figure 2-14: **Examples of networks obtained from transfer entropy.** An directed edge is drawn from one neuron to another, if the directed pair of neurons showed higher transfer entropy than a threshold (which is set to 0.04). The colors of the lines represent modalities of the pre-synaptic nodes: vision: gray, IR: purple, touch: green, reward: red, motor: brown. **HL:** High to low VTE. **L:** Low VTE.

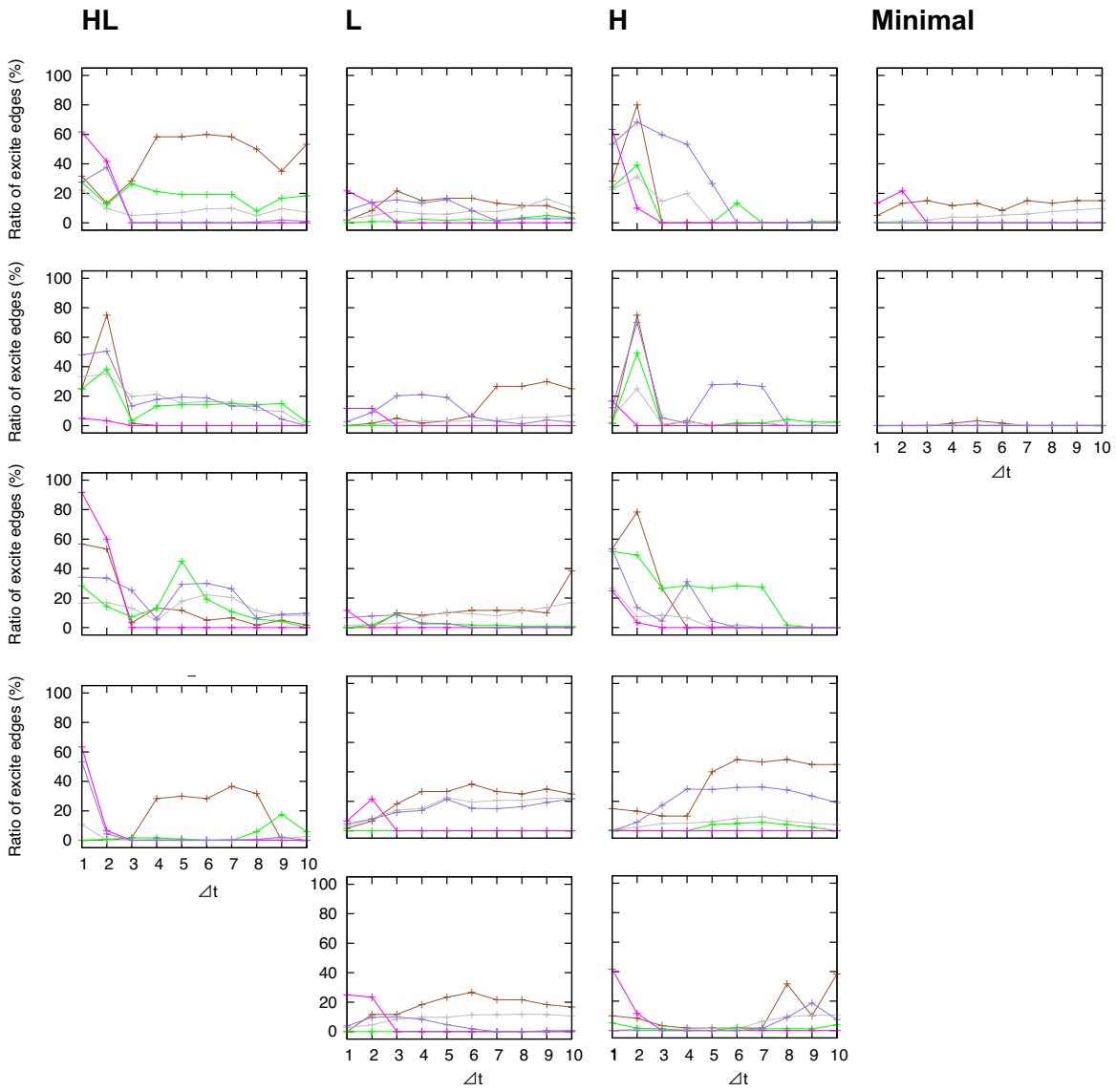

Figure 2-15: **The ratio of edges over all the possible connections obtained from TE.** Each subfigure corresponds to each of 16 evolved robots. When a directed pair of neurons showed higher transfer entropy than a threshold (which is currently set to 0.04; e.g., Figure 2-14), an edge is drawn. The X axis represents time scale Δt, while the Y axis shows ratio of edge over all the possible edges. The colors of the lines represent modalities of the pre-synaptic nodes: i.e., vision: gray, IR: purple, touch: green, reward: red, motor: brown. **HL:** High to low VTE. **L:** Low VTE. **H:** High VTE. **Minimal:** Minimal model.

vation of VTE and the advantages provided by it to demonstrate that VTE is associated with the chaotic activity of neuronal dynamics (Figure 2-5 and 2-7). We modelled VTE in a simulated robotic experiment based on [5], which is a simplified abstract model of learning beings. The aim of this work is not to imitate biological structure of animals, but to construct an abstract model to understand essential mechanism that causes VTE. We expected that this simplified model would show 1) an emergent property of VTE just from a body-environment interaction, 2) the necessary mechanism behind VTE, and 3) the effects of VTE on behavior.

As observed in this work, the spontaneous learning of the correlation between a cue and the location of the goal is achieved after repeatedly exploring the environment. We found that some robots showed conflicting behaviors similar to biological VTE patterns reported in experiments done with rats [108, 111]. A common feature is that the highest frequency of VTE was found at the beginning of the learning stage and gradually diminished after mastering the task (Figure 2-3). Especially at the 6th or 7th trial, which were just after switching the reward position, the robot showed the highest number of VTEs. This could be the results of the interaction between the two neuronal pathways mentioned earlier in the method section; i.e., one path (reward – vision – motor) leads the robot to go to the previous reward position (wrong decision), and the other (tactile – motor) lets it follow the tactile cue (correct decision). Therefore, we concluded that the head movements observed here were similar to real VTE, because they appear to be elicited from a conflict, rather than mere oscillations caused by immature synapses. In the following analyses, we collected more evidences that VTE found in this experiment was not just an epiphenomenon expressed through head oscillations, but actually improved the learning performance of the robot.

We found that with different parameters the robot showed other patterns of VTE, which we classified into three groups: high at the beginning and low afterward (HL), low during the whole learning period (L), and high all the time (H). Those robots belonging to the L and H groups could have different strategies from that explained in the method section. Interestingly, VTE was found only in some neural networks with redundant connectivity, while networks with a minimal ensemble of connections failed to show VTE.

Figure 2-16 summarizes two learning processes with and without VTE. A learning process with VTE (Figure 2-16 left; red line) will go through several steps. First, neurons display chaotic

activity, which creates an exploration motion pattern that allows the sensory inputs to fluctuate and vary constantly. When the chaotic dynamics is maintained, it prevents the robot from falling into a stable attractor. Therefore, the robot repeatedly learns the same environment under different conditions by exploring different paths. This prevents the convergence of the synaptic weights, allowing adaptive control of the behavior (i.e., if an unpredictable event happens, the robot can smoothly adapt to it; Figure 2-12 and 2-13). We believe that these are the advantages of the VTE, as can be seen in Figure 2-9, 2-10 and 2-11). On the other hand, a robot not showing VTE (Figure 2-16 right; blue line) displays a stable activity of its neurons, which lacks the chaotic instability, leading finally to settle into a periodic motion behavior. Such a stable behavior is not conductive to adaptability in general. Therefore, we have verified that the presence of VTE endows the robot with the ongoing control necessary to achieve adaptive behavior. Different from a mere reactive behavior, a robot showing VTE can adjust its behavior to the task, thereby improving its behavior pattern. The ongoing learning found in the VTE-showing robots is also a feature of deliberative decision making, so that this experiments computationally supports that, although it is indirect, a link between them [94, 111].

We can paraphrase the above observation by making the following statement: VTE adds "internal" noise to the system's behavior, so that the performance is automatically geared toward robustness; as a consequence, robots without VTE fail to extrapolate the correct behavior from inexperienced sensory inputs. VTE is associated with the chaotic activity of neurons, where Hebbian learning continuously reorganizes the connectivity patterns, which results in robust behaviors.

The parameters of the robot were evolved through a genetic algorithm, and based on the parameters, the robot had to learn the task by Hebbian learning. Hebbian learning is focused on the autonomous and on-line control of the robot's movements, which we believe to be a biologically plausible learning mechanism. Contrary to a widely used method for controlling robots (e.g., back propagation, which tunes synaptic weights to minimize the error between the expected outputs and its own [92]), Hebbian learning does not require any explicit functions. Due to the lack of the reward function, it is difficult to control robots to solve complex tasks, yet Hebbian learning is an important piece of the mechanisms responsible for the capacity to learn and to memorize information. Although the network structure we used here was specific to the

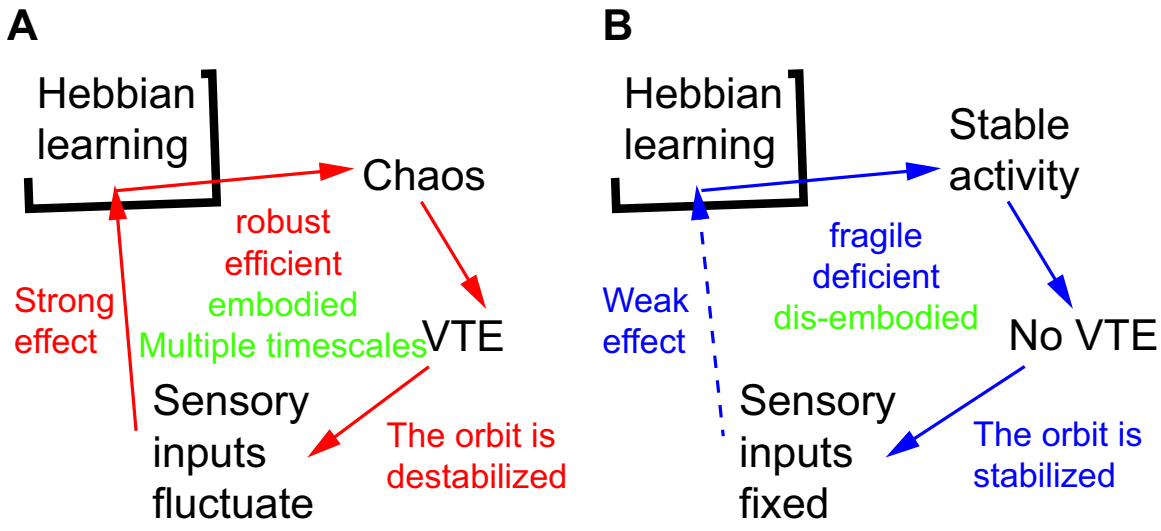

Figure 2-16: **Diagrams showing overall dynamics with and without VTE. A:** The path with VTE exhibits the following. 1) Neurons showed chaotic activity, which destabilises the orbit, making sensory inputs fluctuate and vary constantly. 2) When the chaotic dynamics is maintained, it prevents the robot from falling into a stable attractor. 3) The robot repeatedly learns the same environment under different conditions by exploring different paths, allowing flexible and embodied control of behavior, resulting in robust efficiency. **B:** A robot not showing VTE displays a stable activity of its neurons, leading to a periodic sensory input pattern. This fixed pattern of sensory inputs gives only a weak effect on Hebbian leaning, which results in a fragile reaction to environmental perturbations.

Bovet-Pfeifer model, we think the obtained results do not depend on the detailed architecture of the model. However, we believe that the separation of Hebbian learning from the internal dynamics is an important feature of the model, which is necessary in order to reproduce these results.

Transfer entropy analyses showed that the HL robots had various time scales. This result indicates that having various time scales in a neural network contributes to 1) generation of VTE, and 2) robust control of behavior allowed by VTE.

2.5 Conclusion

To conclude, we believe that linking explorative motions like VTE, chaotic activity, and robust learning is the way to understand how Hebbian learning functions in the real brain system. The connection between robust learning and VTE is consistent with previous research [2,6], in which rats/humans were found to exhibit a better performance when they showed VTE. Perhaps the biologically missing piece is whether chaos is truly part of the mechanisms how the real brain works [26]. We believe that VTE is an essential phenomenon for understanding how the brain works, but, as it is still important to determine whether chaos is an epiphenomenon caused by VTE or the mechanism that causes VTE, the relationship between chaos and VTE must be carefully investigated further in animal experiments (e.g. [52, 94, 111]).

We demonstrated that wide range of time scales were spontaneously organized through learning. In addition, our results presented in this chapter suggest the contribution of various time scales to robust control of behavior. By combining the results from Chapter 3, we discuss it in the concluding chapter Chapter 4.

Chapter 3

Multiple Time Scales Observed in Spontaneously Evolved Neurons on High-density CMOS Electrode Array

Spontaneous evolution of neural cells was recorded around 4–34 days in vitro (DIV) with high-density CMOS micro-electrode array, which enables detailed study of the spatio-temporal activity of cultured neurons. We used the CMOS array to characterize 1) the evolution of activation patterns of each putative neurons, 2) the developmental change in cell-cell interactions, and finally, 3) emergence of multiple timescales for neurons to exchange information with each other. The results revealed not only the topology of the physical connectivity of the neurons but also the functional connectivity of the neurons within different time scales. We finally argued the relationship of the results with "functional networks", which interact with each other to support multiple cognitive functions in the mature human brain.

3.1 Introduction

How can the gap between living and nonliving matter be bridged? Since 1987 when Artificial Life was launched by Christopher Langton, we have not answered this question. Rodney Brooks wrote in the paper (2001) that there are four possibilities why we still cannot make living machines: 1) An alife model is correct, but several parameters were set incorrectly; (2) an alife model needs

more complexity; (3) we need more computational power; and (4) a new fundamental law in addition to the laws of physics is needed. In this paper, we search for the possibility of (2) yet unrevealed laws of neuro-dynamics, by studying cultivated neural cells on a glass plate.

Biological neurons are cultivated on a glass plate from neural "seeds". The seeds develop into either neural or glia cells. Neurons have cell bodies, axons, and numerous dendrites, which the neurons use to connect with each other. An unique characteristic of the present study is that the action potential from neurons was recoded by using a CMOS array glass plate. As we will describe later, each CMOS is the same size order of the neurons. Therefore, by using the CMOS array, it is possible to record the time series of each neural firing. A remarkable aspect of this biological neural network is the developmental process. The entire time course of the growing process can be recorded with the CMOS array. We analyze the time series data to characterize the developmental dynamics.

A disadvantage of this experiment is that there is no way of designating which neurons connect to which. We thus measure the information transfer from the time series to infer the neural connectivity. This method reveals not only the topology of the physical connectivity of the neurons but also the functional connectivity of the neurons within different time scales. A finding here is that growing biological neurons use different time scales to exchange information with each other.

This chapter is organized as follows. In Section 3.2, the specifications of the CMOS array and the biological conditions for the neural cells are provided. The method for cultivating cells and associated techniques are also described. In Section 3.3, the analyzed results are presented. The activation patterns of each cell are quantified with inter-spike intervals (ISIs). Cell-cell interaction is also inferred by transfer entropy (TE), which reveals that multiple timescales emerge from the neural network. Finally, in Section 3.4, this chapter is summarized.

3.2 Materials and Methods

To measure the electrical activity of cultured neurons, we used a high-density CMOS micro-electrode array [1] [33]. The CMOS array is pictured in Figure 3-1A. This array is an emerging

[1]All procedures were approved by the institutional committee at the University of Tokyo, and were performed in accordance with the Guiding Principles for the Care and Use of Animals in the Field of Physiological Science of

instrument for investigating the spatio-temporal activity of cultured neurons in detail. The CMOS array has 11,011 recording sites with an inter-electrode distance of 18 μm, i.e., in the order of cell body size, and a sampling rate of 20 kHz. This high spatio-temporal resolution allows precise recording of action potentials from the identified cell bodies of neurons. Using this high spatial resolution, neural somata are localized and their activity is recorded.

This CMOS array is superior to conventional microelectrode array (MEA) [29, 105, 112] in respect of spatio-temporal resolution: The locations of recording sites in conventional MEAs are predetermined, with an inter-electrode distance of 200 μm, so that it is difficult to identify signals from an individual cell, and neurons far from these recording sites are not included. Alternatively, optical imaging can be used to study the Ca^{++} dynamics of any neuron of interest; however, the temporal resolution is not high enough to characterize the action potentials of each neuron.

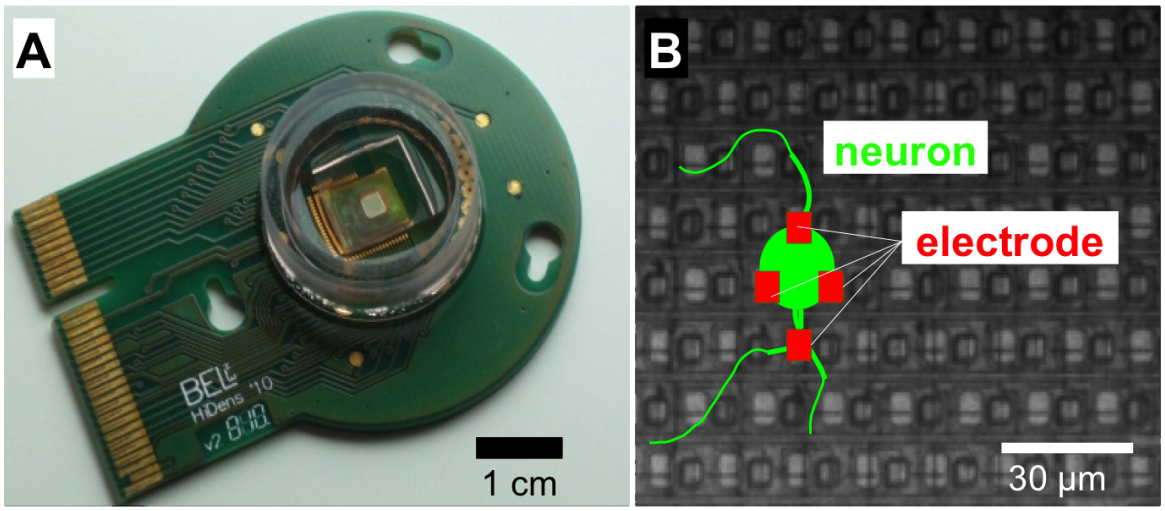

Figure 3-1: **The High-density CMOS array.** A: Appearance. B: A picture to show the configuration of electrodes. Interval of each electrodes is close to the cell body size.

Dissociated neural culture The neural cultures were prepared from the cerebral cortex of E18 (embryonic 18) Wistar rats. The cortex was triturated with trypsin and dissociated cells were plated and cultured on high-density CMOS microelectrode arrays coated with polyethylenimine and laminin. For the first 24 h, the cells were cultured in neurobasal medium containing 10% horse serum, 0.5 mM GlutaMAX and 2% B-27 supplement. After

the Japanese Physiological Society.

the first 24 h, half of the medium was replaced with growth medium in the form of Dulbecco's modified Eagle's medium with 10% horse serum, 0.5 mM GlutaMAX, and 1 mM sodium pyruvate. During the cell culturing, half of the medium was replaced once a week with the growth medium. The cultures were placed in an incubator at 37 °C with an H_2O-saturated atmosphere consisting of 95% air and 5% CO_2. We prepared for two conditions with different densities of neurons. The denser chips had 35,000 cells plated, denoted as Chip#1D and Chip#2D, while the sparser ones had 14,000 cells, termed Chip#1S and Chip#2S. The number of the active neural somata was around 100, which was estimated by the array scanning (detailed in the next part). Although the reason is not clear, the other cells, which were not detected as somata, may be either inactive cells, glia cells, or dead cells.

Recording of neural activity Neural activity was recorded with high-density CMOS microelectrode arrays. Before the neural somata activity was recorded, almost all of the 11,011 electrodes were scanned to obtain an electrical activity map with which we estimated the locations of the somata. A scan session consisted of 95 recordings; each recording was conducted for 60 sec with about 110 electrodes that were selected randomly, while avoiding overlap with already selected electrodes. An electrical activity map was obtained from the scanned data by calculating the average height of the spikes for every electrode. It was assumed that the neural somata were near the local peaks in the Gaussian-filtered electrical activity map. About 100 of the higher-level peaks were selected, and then the nearest electrodes were selected for recording. If the number of local peaks was fewer than 126, then all the peaks were selected. The electrical activity of the selected electrodes was recorded for 30 min. All recordings were done at a 20-kHz sampling rate using the LimAda spike detection algorithm [116] with a threshold of 5.0. Unexpected double-detected spikes were removed from the data before the analysis was conducted.

3.3 Results and Discussion

3.3.1 Simple observation of activity patterns

First, we observed the activation patterns of the neurons by examining the time series of the neural spikes. Figure 3-2 shows examples of raster plots of Chip#1D and Chip#1S. Chip#2D and Chip#2S showed similar tendency to their respective counterparts (data not shown). The data plotted here are a compressed version of the raw data with a different bin-length, which is denoted by Δt. Namely, spikes within the same bin are regarded as one spike. If Δt is smaller, the time series represents single spikes generated from each single neuron, while a larger Δt represents the macroscopic behavior created by an ensemble of spikes. In Figure 3-2, the data are plotted in $\Delta t = 0.6$ ms and 100 ms. At the top of each raster plot, the activation ratio is displayed. We chose $\Delta t = 0.6$ ms and 100 ms as the examples by two reasons; a single spike of neurons lasts around 1 ms, while, observably from Figure 3-2, synchronous activation of neurons is detected around Δt =10–100 ms. Therefore, we took $\Delta t = 0.6$ ms to capture interaction between individual cells (microscopic), while $\Delta t = 100$ ms is to observe collective activities generated from an ensemble of neurons (macroscopic).

As shown in Figure 3-2A right, the Chip#1D neurons at days *in vitro* 7 (DIV 7) are activated intermittently in synchronization. This activation pattern was burst synchronization, in which an ensemble of neurons has active and silent phases; in the active phase, synchronous activation of the neurons is intermittently observed. The burst synchronization is a typical activation pattern observed in cell cultures [67]. On the left of Figure 3-2A, the silent phase of the neural activation is magnified with $\Delta t = 0.6$ ms. Neural spikes are observed sparsely, which suggests that each neurons was activated independently from others.

On the other hand, at the later stage (DIV 14) of the Chip#1D, burst synchronization is not observed (Figure 3-2 B right). Instead, we observed each neurons shows different activation patterns; i.e., some neurons were activated almost all the time, some showed spikes less often, and others remained silent. These neurons seem to be activated at different frequencies. The left of Figure 3-2B shows that, at $\Delta t = 0.6$ ms, a single spike is followed by another, which suggests that one single spike can activates another. In contrast, Chip#1S showed the burst synchronization throughout the entile recording period (Figure 3-2C and D).

To summarize the results thus far, burst synchronization was observed in two conditions; in the earlier developmental stage of the dense cell condition, or in the sparse cell condition. In those cases, any single spikes hardly activate another. In the later stage of the dense cell condition, the neurons spiked at various frequencies, where single spikes induce others.This tendency is explained by the maturity of synapses. At the earlier developmental stage, the synapses are not mature; therefore, a single spike cannot activate another. Still, if the neurons send spikes at the same time (synchronous bursts), the signals become enough strong to activate each other. In contrast, at the later stage of development, the synapses were enough strong to transmit a single spike from one neuron to another.

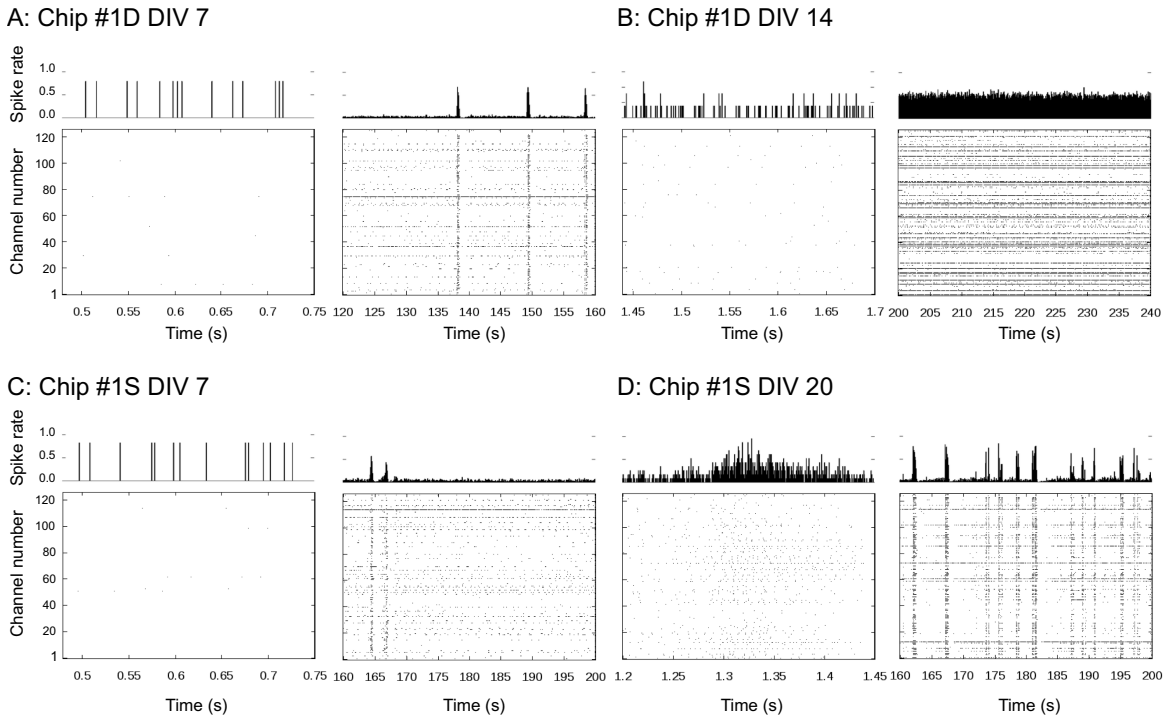

Figure 3-2: **Examples of the raster plot of the cultured neurons.** The X axis denotes time (s). The Y axis represents the indexes of the recording channel, where one channel can be considered as one neural cell. The subfigures at the top of the raster plots shows the spike rate. **A–D:** Results for Chip#1D and Chip#1S with different DIVs. $\Delta t = 0.6$ ms (microscopic) and 100 ms (macroscopic).

3.3.2 Single cell activity evaluated with ISI

In the next step of the analysis, we quantified the activation patterns of each neuron. Therefore, we analyzed the inter-spike intervals (ISI) distribution of the neural activity. Figure 3-3A depicts examples of ISI distribution recorded from putative single neurons cultured in Chip#1D. In the earlier culture stage (DIV 7; Figure 3-3A left), exponential decay was observed, while the neurons tended to obey the power law at DIV 14 (right figure). To quantify this tendency, we plotted the ISI frequency on a logarithmic scale, and we fit the slope with a straight line by the least squared method. If the ISI distribution follows the power law, the distribution should be a straight line, so that the R-squared value is an index of fitting the power law.

Figure 3-3B shows the average of the R-squared values over each neuron; the values tended to increase in Chip#1D, Chip#1S and Chip#2S. Especially, Chip#1D shows a drastic increase at DIV 10, which suggests the neural activity became closer to the power law. Chip#2D also shows a higher R-squared value throughout the entire recording period. The recordings for Chip#2D started from DIV 17, which explains the result is similar to the later stage of Chip#1D.

The power law means that each neuron exhibited a wide range of ISIs. It may be related to the observation from the raster plots (Figure 3-2), which showed that a wide range of frequencies between neurons were observed at the later development stage. To sum them up, a broader range of time scales likely emerges after synaptic maturation. However, from this ISI analysis, it is not possible to understand how those cultured neurons interact with each other to generate the activity patterns. In the next part, we then investigated neuron connectivity.

3.3.3 Cell-cell interaction inferred with transfer entropy

We used transfer entropy (TE) to estimate the effective connectivity for transferring information from one neuron to another. TE measures directed information transfer, which detects causal relationship between two time series [2, 65, 95, 103]. For instance, higher TE from one neuron to another indicates that the first neuron strongly affects the second. Therefore, TE enables to find the functional synaptic connectivity. The specific definition of TE is presented in Section . We applied it to artificial neural networks to ensure the validity of TE to estimate effective connectivity. Then, we applied TE to the cultured neural cells to infer their topology.

Settings of artificial neural networks TE analysis was first applied to a computational neural model. The model was built around Izhikevich neurons connected through artificial synapses [51]. The Izhikevich neurons form a simple model of cortical neurons that is implemented by a system of two differential equations modeling the membrane potential and the refractory period. When the membrane potential reaches a threshold value (for instance, 30 mV), a spike is emitted. This spike is transferred to post-synaptic neurons through some shared synapses. The voltage on arrival is the original spike strength, modulated by the efficacy of the synapses. For instance, an initial spike of 30 mV traveling on a synapse with an efficacy of 0.5 delivers a voltage of 15 mV to the post-synaptic neuron. Every synapse has a delay of 1 ms between the time of emission and the arrival of a spike.

The complete model is composed of seven neurons: four input neurons receiving randomly generated external stimulations, two internal neurons and one output neuron. The parameters for the Izhikevich neurons correspond to the regular spiking model($a = 0.02$, $b = 0.2$, $c = -65mV$ and $d = 6$). Different types of connectivity patterns have been tested, ranging from fully interconnected to sparse (Figure 3-4A–C). The strength of the connection is randomly assigned based on uniform distribution. Every update of the model represents a 0.1-ms step in time, which ensures the model's mathematical stability. The total duration of a test is 1000 s, which corresponds to 10,000,000 updates of the model.

Estimated connectivity of the artificial network From the time series of the artificial neural activity, we calculated the TE from one neuron to another. Using the TE, we estimated the network structure of the artificial neurons. A synaptic connection from one to the other was assumed when the TE in a directed pair of neurons was higher than a threshold, which was currently set to 0.05. Then, we compared the topology of the reconstructed network and the original network shown in Figure 3-4A–C. The number of false edges for each topology is shown in Figure 3-5. There are two types of false edges: one is false positive, and the other is false negative. False positive means the edge that is estimated to exist, but actually not, while false negative represents the edge that exist in the original topology, but is estimated to be there. Figure 3-5 shows the sum of the false negative and positive. Δt in this figure is the same one as previously used, which represents the

bin-length of the compressed time series. This figure shows that the optimal parameter set to reconstruct the original topology depends on the dimension of the past variables (k and l) and Δt.

The result ensures that the effective connectivity to transmit signals by TE is estimated. However, a good approximation depends on the dimension and time scale Δt. We used various Δt to get a good approximation of the effective connectivity in the cultured neurons.

Estimated connectivity of cultured neurons Figure 3-6 shows some examples of the estimated network structure of neurons. This depicts the network of Chip #1D DIV 14 with different Δt ($= 0.6$ ms, 1 ms and 10 ms). An edge is drawn if the TE from one neuron to another is higher than the threshold, which was set to 0.00001. The threshold is determined arbitrarily, but to display dynamical change in connectivity patterns. The first observation about this figure is that different topology is structured depending on Δt.

Development of functional connectivity Figure 3-7 shows network structure obtained with $\Delta t = 0.6$ms and 10ms through development. The upper figures for each chip correspond to $\Delta t = 0.6$ms and the lower ones denote $\Delta t = 10$ms. This figure shows that functional connectivity changed over time, where different network structures were observed with respective time scales.

As shown with the artificial neural network, each connection may not have the same optimal Δt to estimate connections. Based on [79], we used the optimal Δt to understand the information flow in the network.

3.3.4 Optimal time scale to convey information

As observed, the cell-cell interaction estimated from the TE analyses depends on the time scale Δt, so that we evaluated the optimal Δt to estimate the network structure better. We defined the optimal Δt as Δt^* that maximizes TE. Δt^* was calculated for each directed pair of neurons. Information was considered to be transferred most effectively with the Δt^*. Therefore, the Δt^* exhibits effective synaptic connectivity from each neuron to another.

Figure 3-7A and B depict Δt^* from one neuron to another in Chip#1D and Chip#1S respectively. Edges are drawn if TE at the Δt^* is larger than the threshold ($= 0.00001$; same as the one used above). The color indicates the value of Δt^*, where red represents Δt^* smaller than 10 ms and blue indicates larger. We defined $\Delta t^* = 10$ ms as a threshold to assign color, because burst activity was observably detected with Δt larger than 10 ms (Figure 3-2). Hereafter, we call $\Delta t^* < 10$ ms "the smaller Δt^*", and $\Delta t^* \geq 10$ ms "the larger Δt^*", where we divide the activity pattern into microscopic and macroscopic using this threshold.

Interestingly, the smaller Δt^* coexisted with the larger Δt^* (Figure 3-8A and B). In addition, this figure shows that there were several intervening neurons that have both red and blue edges. This implies that microscopic and macroscopic activity can exchange information with each other.

Figure 3-9 shows the Δt^* distribution. Several tendencies were shared by all of the chips: 1) As DIV increased, Δt^* became distributed in a wider range, or more specifically, smaller Δt^* than 10ms became observed more often after DIV 12 for the dense condition (Chip#1D and #2D) and after DIV 16 for the sparse condition (Chip#1S and #2S). This suggests that causal relationships between neurons became happening in a shorter period. This is consistent with the observation in Figure 3-2B that a single spike was followed by other spikes within a shorter period. Additionally, 2) the number of peaks in the frequency of Δt^* became high compared to the earlier stage of the development (i.e., DIV 7–8 as shown in Figure 3-9). Namely, the frequency of Δt^* varied. From those observations, it is plausible to say that information became transferred at multiple time scales in a wider range through development.

3.4 Conclusion

We characterized neural activity cultured on a high-density CMOS array. The neurons showed different activation patterns depending on the density and age (days *in vitro*; DIV). The neurons did not receive external inputs, but still spontaneously evolved. The neurons were cultured in two different conditions, i.e., dense or sparse.

When the cell density was high, the neurons showed burst activity first, and afterward, each neural cell activated at a different frequency (Figure 3-2A, B). Some were activated all the time,

some showed spikes intermittently, and others remained silent. The ISI distribution of each cell became closer to the power law as the DIV increased (Figure 3-3), which suggests a single cell evolved to show a wider range of time scales. These results suggest that neural activity exhibits a wider range of time scales as the synapses mature. This tendency was shown clearly with TE (Figure 3-7). At the earlier stage of development, TE was mainly maximized at a larger time scale, while, at the later stage, TE was also maximized at a smaller time scale (Figure 3-9).

A mature human brain is a collection of functional networks, each of which corresponds to a different cognitive function [30]. In this chapter, we insist that even a neural network on a glass plate spontaneously develops multiple “functional networks”, which can be distinguished in terms of the time scale determined by effective information transfer. Without relevant sensory input, we cannot say the networks are functional in the proper context of brain science; however, we speculate that the spontaneous development of “functional networks” is a candidate for the brain functional network.

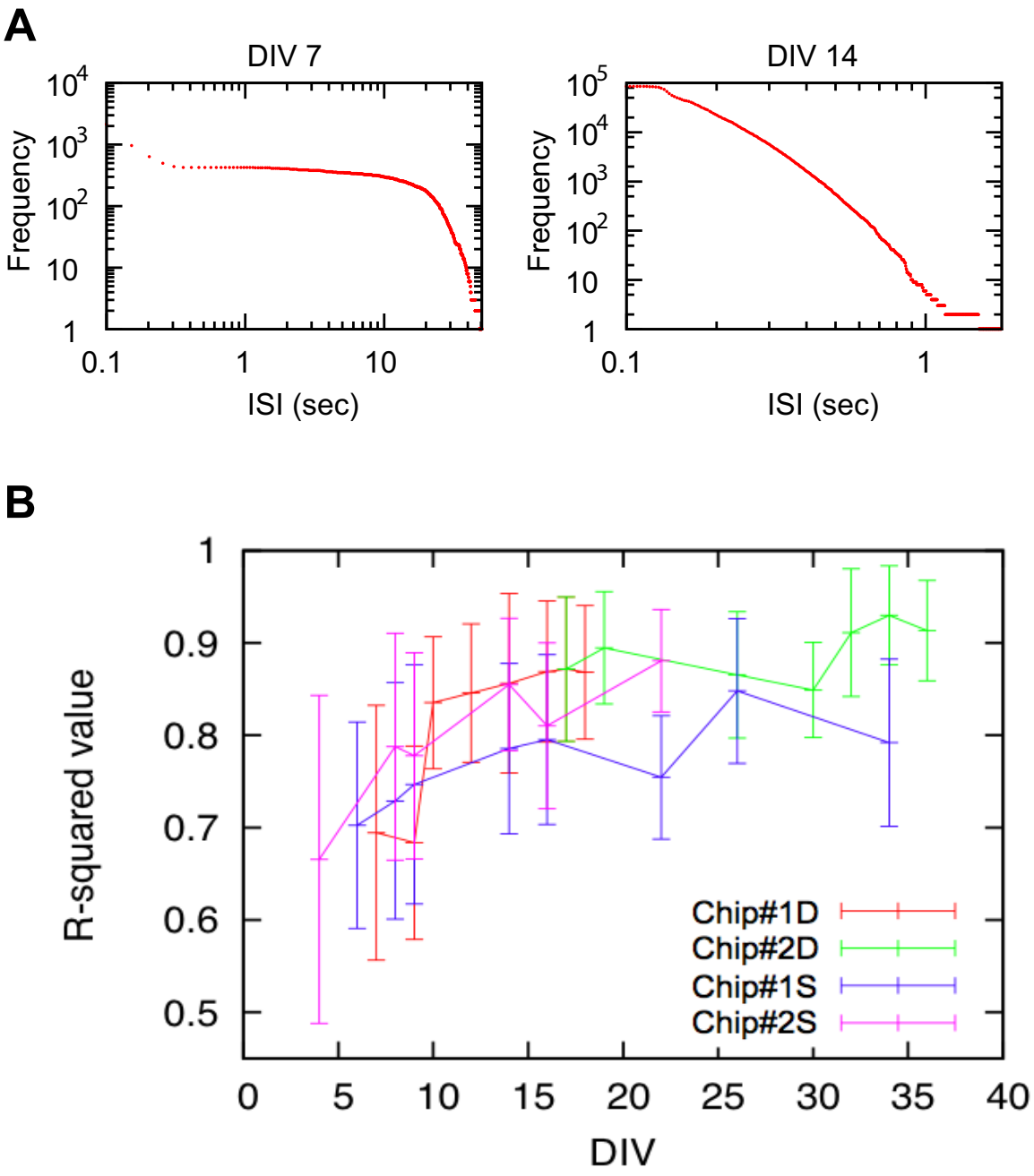

Figure 3-3: **ISI distribution of each neuronal cells. A:** Examples of the ISI distribution of neural activity recorded in a cell on Chip#1D. The X and Y axes depict the ISI logarithm and frequency respectively. The two figures show the results for each 7 and 14 DIV. The estimated exponents for DIV 7 and 14 are -1.40(0.57) and -3.77(0.57). The values inside the parentheses denote the R-squared values of the regression lines. **B:** Change in the R-squared values with DIV. The R-squared value is obtained from a regression line fit to the ISI distribution, which is plotted on a logarithmic scale. When the R-squared value is closer to 1.0, the ISI distribution is more likely to obey the power law.

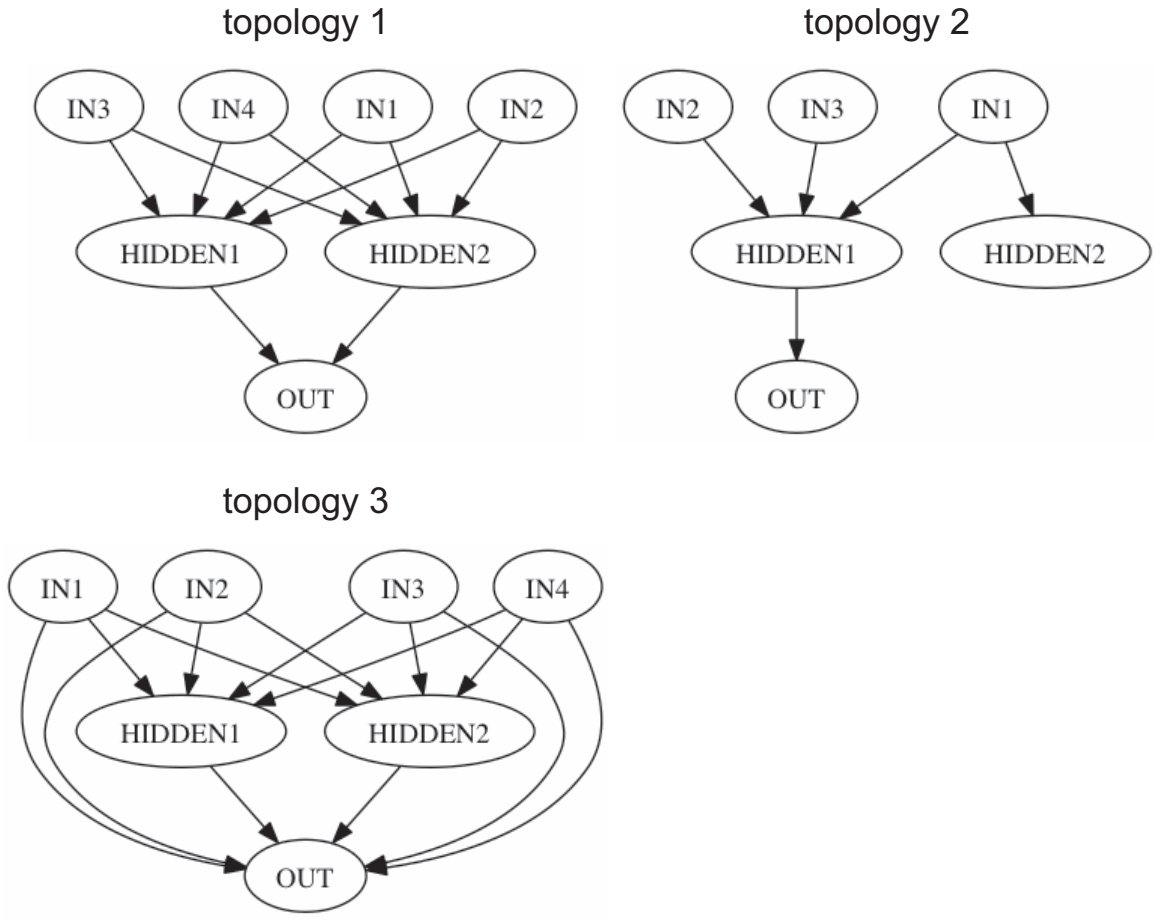

Figure 3-4: **Topology of the original network used in the computer simulation.** Three connection patterns were used.

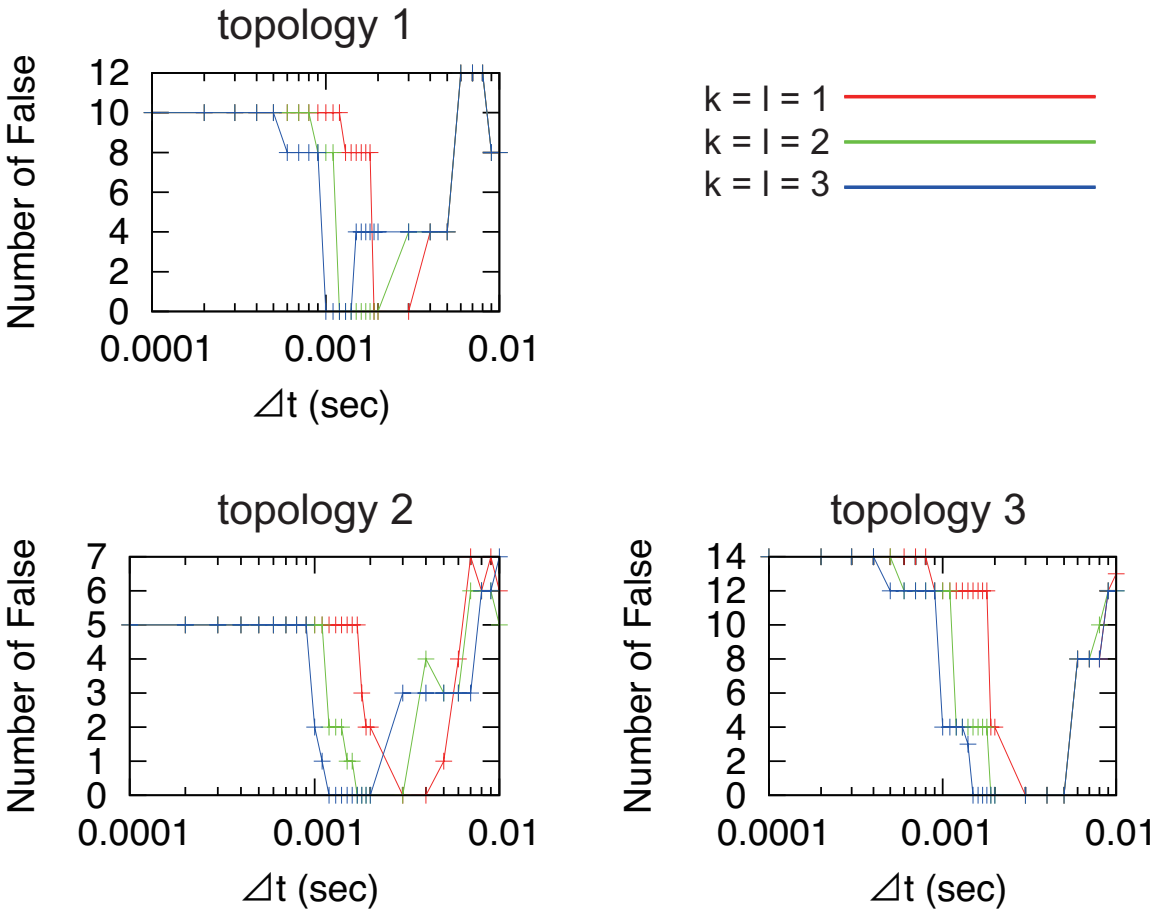

Figure 3-5: **Number of false connections with different Δt values.** This figure shows the sum of the false positive / negative edges. The threshold was set to 0.005.

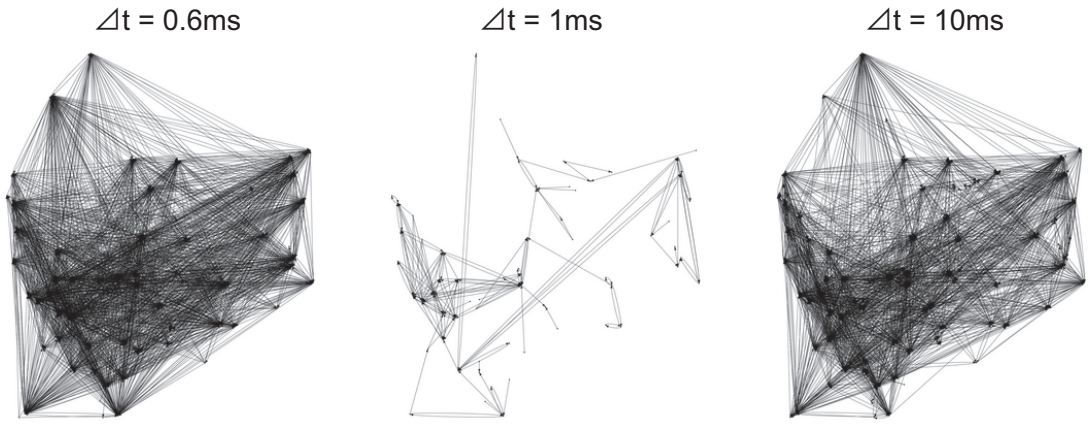

Figure 3-6: **Examples of the network structure of the information flow created by TE.** The data used here is from Chip#1D DIV 14. $\Delta t = 0.6$ ms, 1 ms, and 10 ms. The TE threshold to draw edges was set to 0.00001.

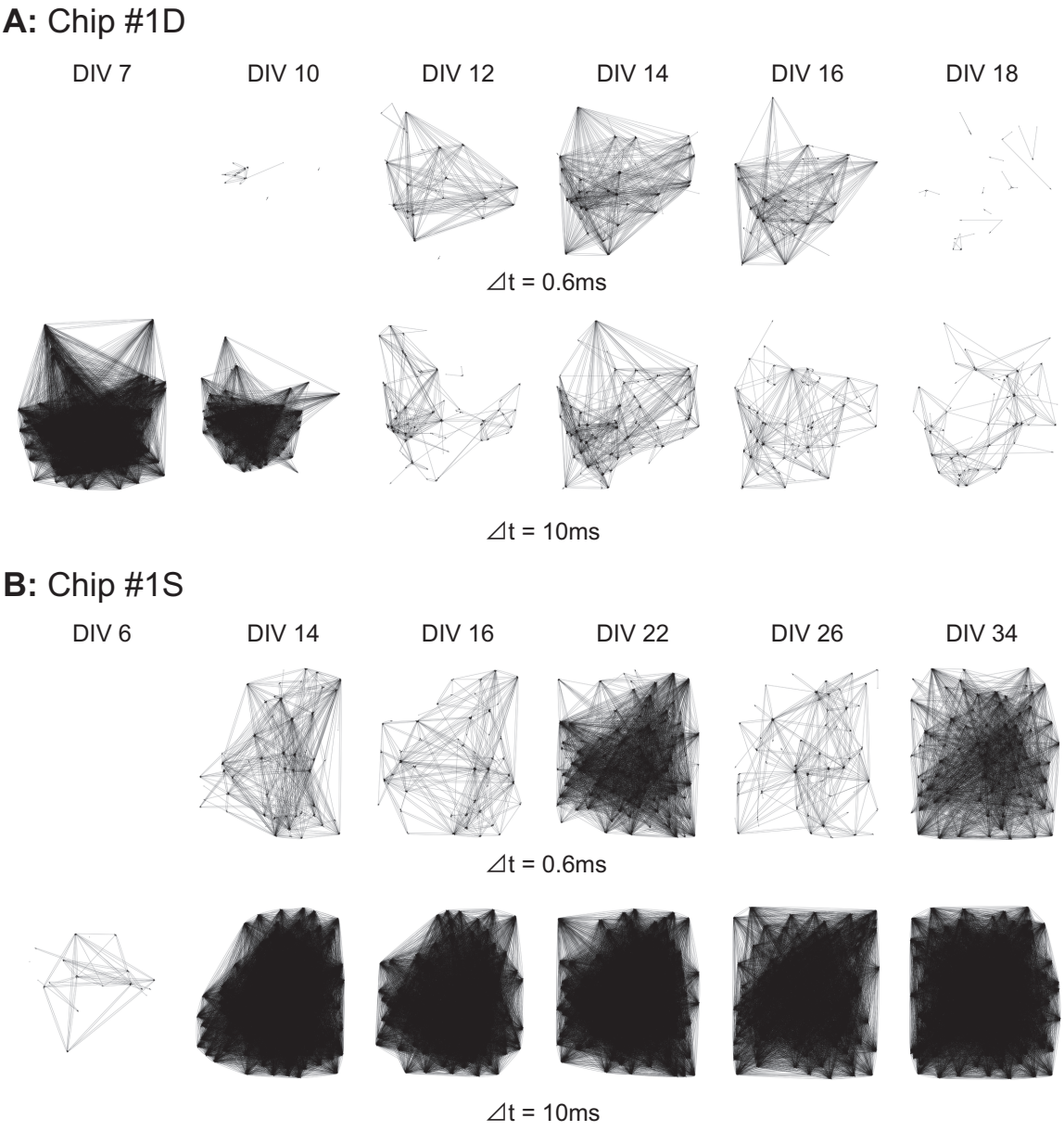

Figure 3-7: **Network structure obtained with $\Delta t = 0.6$ms and 10ms. A:** Results for Chip #1D. **B:** Results for Chip #1S. The upper 6 figures in each **A** and **B** correspond to $\Delta t = 0.6$ms and the lower ones denote $\Delta t = 10$ms. The TE threshold for drawing edges is equal to 0.00005.

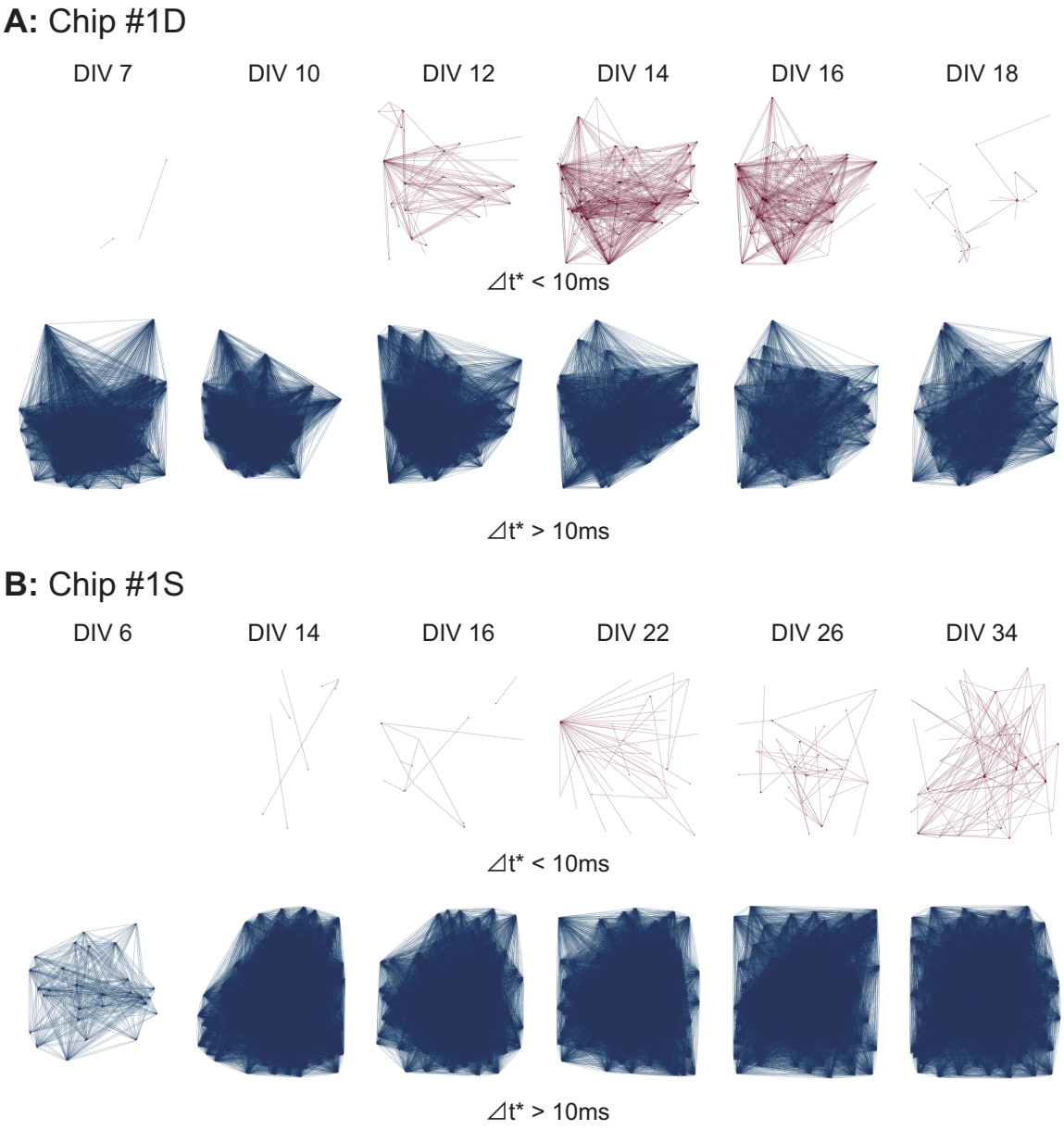

Figure 3-8: **Examples of network structures obtained with Δt^*. A:** Results for Chip #1D. **B:** Results for Chip #1S. The color indicates the value of Δt^*, where red represents Δt^* smaller than 10 ms, and blue is for a larger Δt^*. The TE threshold for drawing edges is equal to 0.00001.

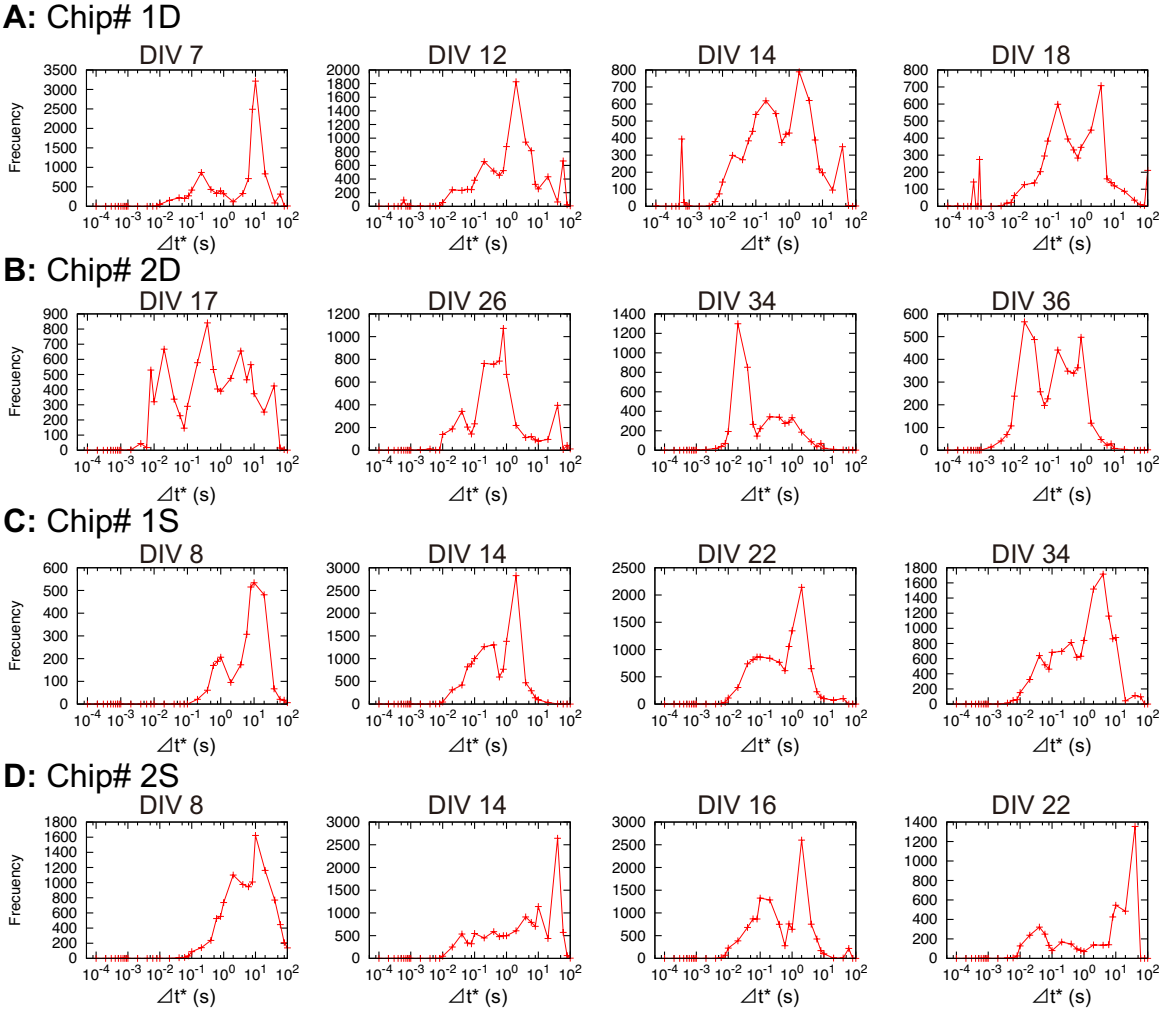

Figure 3-9: **Distribution of Δt^* in accordance with DIVs.** The X axis denotes Δt^*, while the Y axis shows the frequency of Δt^*. **A:** Chip#1D, **B:** Chip#2D, **C:** Chip#1S, and **D:** Chip#2S.

Chapter 4

General Discussion

This concluding chapter summarizes the results obtained throughout the thesis, discusses the general principles that can be drawn from the behaviors observed in the preceding chapters, and considers future perspective of the proposed approach to the study of adaptive behavior.

4.1 Summary

As mentioned in the introduction chapter, the contribution of this thesis lies in exploring the way to consist life-like phenomena from self-organizing processes, especially focusing on time scales. Based on the synthetic methodology of the constructive approach, I first build artificial systems, then observed structures self-organized by synaptic plasticity, and finally analyze them. I constructed two different systems, i.e., robotic experiments, and biological neural systems, where I observed adaptive and complex patterns that worked on different time scales, thorough a course of spontaneous development or learning. The following is the summary of the results obtained from Chapter 2 and 3.

Chapter 2 presented a series of experiments with a multi modal robot, highlighting behavioral change during a course of learning, i.e., change in vicarious trial-and-error (VTE). VTE have been studied mainly in route selection tasks with rats, where they shook their head from one option to another, as if they were considering their decision between the two conflicting options. The number of VTE (one VTE was counted when the rat looked at one option and then it looked at the other option) was high at the beginning of learning, and became less afterward. More recent

studies suggested that VTE was observed when the rats were uncertain in making decision, so that it is said to be relative to deliberative decision making, i.e., searching, predicting and evaluating potential outcomes [111], which allows flexible and ongoing control of behavior.

In the robotic experiment, I observed a similar VTE pattern to the rats, i.e., higher at the beginning of learning and less afterward, which I named the "HL" pattern. With different parameters, the robot sometimes showed the HL pattern, and sometimes not (i.e., the amount of VTE was kept either less or high during the whole trials). Regardless of the VTE patterns, the robot was still successful in solving a given task, so that VTE appeared to be generated from redundant mechanisms in the neural network. In fact, VTE was found only in some neural networks with redundant connectivity across modalities, while networks with a minimal ensemble of connections failed to show VTE.

Although VTE seemed to be generated from a redundant mechanism, it was functionally not redundant, but allowed robust control toward perturbations. Even when perturbations were given to the environment, the HL robot kept the success rate at a certain level, while the other types of robots easily dropped it. This robust behavior was explained by unstable activity in the neural network. The maximum Lyapunov exponent (MLE), which is an index of chaos, showed chaotic neural activity in the VTE-showing robot, and stable activity in the other robots. The chaotic activity created an exploratory motion pattern that allowed the sensory inputs to fluctuate and vary constantly, where the robot repeatedly learned the same environment under different conditions. This prevented the convergence of the synaptic weights, allowing robust control of the behavior. Namely, even when an unpredictable event happened, the robot was able to smoothly adapt to it. I concluded that the robust control was the advantages of the VTE. Additionally, computation of transfer entropy showed that the neural activity of the HL robots had multiple time scales in information flow, which is discussed later by relating it with the robust control.

Chapter 3 provided a series of experiments where spontaneous development of neural cells was recorded around 4–34 days *in vitro* (DIV). For the recoding, I used the CMOS micro electrode array, which enables detailed study of spatio-temporal activity of cultured neurons. Indeed, the temporal and spacial high resolution of the CMOS array allowed recordings of single spikes from each putative neurons, which means quite high resolution compared to conventional mem-

brane electrode arrays [33]. I observed that neurons exhibited different activity patterns through development. Neurons exhibited burst synchronization at the earlier stage, and later showed seemingly complex patterns. This observation of the neural activity was supported by distribution of inter-spike-intervals (ISI). At the earlier stage of the development, most of the neurons showed ISI obeying exponential decay, while, at the later stage, many neurons exhibited power law distributions. The power law suggested that neural activity exhibited a wider range of time scales. This tendency was shown clearly with transfer entropy (TE). At the earlier stage of development, TE was mainly maximized at larger time scales, which seemed to result from the burst synchronization observed in the raster plots. On the other hand, at the later stage, TE had its peak around smaller time scales as well as larger ones, which supports the existence of a wide range of time scales observed both in the raster plots and in the ISI distribution. The larger time scales were around 10–1000ms, which appear to result from multi-cellular interactions. By contrast, the smaller time scales were around 1ms, suggesting the existence of single cell interaction. This can be explained by maturation of synapses. At the beginning of the development, synapses were immature, so that only ensembles of neural spikes can activate other neurons. Maturation of synapses allowed each neuron to transmit signals to another, which result in TE with the smaller time scale.

I demonstrated in this experiment (Chapter 3) that neurons *in vitro* can exhibit multiple time scales. In general, the mature human brain is a collection of functional networks, each of which corresponds to a different cognitive function. I consider the network with multiple time scales which we observed in the thesis can be a candidate for the brain networks to represent different functions. If this is true, then multiple functionality will be an innate property of the neuron. This is discussed below.

4.2 Time Scales

Throughout the works of this thesis, I computed transfer entropy (TE), to estimate information flow on the neural network. I observed in Chapter 3 that cultivated neurons spontaneously developed a wide range of time scales that were distinguished by TE. Chapter 2 demonstrated that the VTE-showing robot had various time scales in the neural network. In this section, by

combining those two research works, I discuss the meaning of the time scales that are distributed widely.

4.2.1 Functionality of the Multiple Time Scales

In this part, I consider the meaning of networks distinguished by time scales. Although it is not evaluated, I speculate that each network corresponds to a certain function. As presented in Chapter 1, regulation of the time scale is critical for maintaining functions, and respective functions in turn work on specific time scales. In this respect, it is possible to hypothesize that networks distinguished by the time scale can work as different functions.

The VTE-showing robot (Chapter 2) indirectly supports this idea: Multiple time scales were observed with conflicts, i.e., VTE. Therefore, it is plausible to say that the multiple time scales contributed to generation of VTE. In other words, the multiple time scales allowed to have conflicting different functions. This result suggest that multiple functionality emerges when various time scales are organized in the brain system.

The neural networks in Chapter 3 cannot be functional in the proper context of the neuroscience, because the cultivated neurons did not have any sensory inputs nor motor outputs to demonstrate their functions. I consider that the problem will be solved by connecting the biological neural network with a robot body. This “biological” brain robot will allow to assess functionality of the neural network, and will evaluate meanings of the network distinguished by time scales.

By synthesizing the works in Chapter 2 and 3, it is possible to say that the robot with the biological brain will be a promising and important experiment to shed a new light on the field of ALife. The robot will develop a wide range of time scales in the neural network as presented in Chapter 3, which can allow to obtain various functions to generate VTE as shown in Chapter 2. VTE is an behavioral indicator of deliberative decision making, i.e., prediction and evaluation of possible outcomes [111], which might be a primitive form of conscious thinking. Therefore, the robot will give deeper understandings for adaptive behavior, and conscious thinking.

4.2.2 Robustness

As presented in the introductory chapter (Section 1.1.1), fMRI studies suggest that robust control can be achieved by having multiple functional networks [24, 25]. The computational model presented in Chapter 2 supports their suggestion: The HL robots, which had neural activities working on various time scales, acted in a robust and flexible way. However, the 'H' robots (i.e., the robot showing VTE during the whole period of learning) also showed multiple time scales of TE, but they were not robust to perturbations. This implies that variation in time scales may not be a sufficient condition to generate robust behavior. Still, it is possible to consider that having multiple time scales is a necessary condition to achieve robust control, if, as suggested above, networks specified by time scales can work as specific functions to cause conflict, resulting in deliberative decision making.

In general, from the clinical approach, it is difficult to demonstrate effects of neuronal activity on behavioral levels, e.g., flexibility and robustness, while, as shown here, ALife approaches allow mechanical understandings by creating simplified and abstract models of life. In turn, clinical findings, as reviewed above, give concrete examples to validate the theoretical counterparts. As such, the theoretical and experimental approaches should relay on each other to generate deeper understandings of living systems. In this respect, the outcome demonstrated in this thesis will become a new and strong example in the field of ALife.

4.3 Synaptic plasticity

The two works presented in this thesis both have synaptic plasticity. In this section, I discuss the effect of learning. First, I review conventional learning algorithms, and then compare them with the learning of the work of this thesis, to extract a conclusion.

Learning algorithms in artificial neural networks are classified into two: supervised and unsupervised learning [82]. In supervised learning, a desired state is defined to compare actual outcomes of a network. The error between the desired state and current outcome becomes a feedback to modify the synaptic connections in the network [19, 81]. By repeating the small modifications of synapses, the network achieves to generate desired outputs in response to input signals. Backpropagation and bayesian inference are examples of supervised learning [91, 92].

Unsupervised learning does not have such teaching signals. However, most of the applications of it have target patterns that work like desired states. An example of the application of unsupervised learning is a self-organized map (SOM) [55], which maps input signals onto a set of output responses without changing topological order. The SOM reduces high dimensionality of the input signals into low-dimensional visible patterns. Hebbian learning can be another example of unsupervised learning [16], because it realizes principle component analysis, which is another powerful tool to visualize large data sets [38]. The motivation of studies of unsupervised learning normally lies in pattern recognition, so that data sets that unsupervised learning is applied to are fixed and given ones.

The model presented in the thesis was different from the conventional applications of learning algorithms, in respect that 1) the learning did not need any evaluation values, e.g., fitness functions nor target patterns, and 2) the system was open to the environment, so that input data sets changed continuously depending on the interaction between the environment and body. In other words, the systems in the thesis did not have any explicit motivational drives to give equilibrium points to the systems. Indeed, the systems in this thesis remained unstable: the robot in Chapter 2 included chaotic activity in its neural network, and the robot changed reaction patterns easily by perturbations, while the neurons *in vitro* (Chapter 3) remained unstable to change the structure. Even though they seemed to be at a unstable and transient phase, the systems formed meaningful structures, i.e., VTE and reward seeking behaviors (Chapter 2), and multiple time scales (Chapter 2, 3).

Then what was the driving force to lead the systems to those meaningful structures? I suggest that the driving force to generate them is a self-organized structure from complex interaction in sub-systems. The sub-systems can be decomposed into several layers with having different time scales, e.g., sensory inputs, synaptic plasticity, and behavior. Those are not separated but interacting with each other. Mathematically put, 1) activity of neurons is corresponding to a state vector, 2) synaptic plasticity is parallel to a parameter set, and 3) behavior, that tune synapses, can be represented as meta-parameters. Namely, I suggest that organization of multiple time scales can act as hierarchical orders from the state space to the meta-parameter space, and that will allow generation of adaptive behavior.

Indeed, a brain model with the hierarchical order has been presented, called hierarchical

organized generative models (HGMs) [34,36]. This model is structured by the states, parameters, and meta-parameters. Because of the hierarchical structure, HGMs have demonstrated strong capability to predict various sensory signals. The HGM is often presented as a representation of a theory of the brain called "predictive coding". In this theory, perceptual contents are treated as products of prediction about causes of sensory signals [17,96]. It allows simple and seemingly natural understandings of perception, because it does not require any motivational drives like fitness functions, but still explains various cognitive phenomena, such as attention, hallucination, or synesthesia [32,96]. Although some agent based models have been presented [35], they are yet limited to stable and close systems.

The work of this thesis supports the possibility that the hierarchical order creates adaptive and coherent behavior in open and dynamical systems. Of course, our systems did not assume minimizing predictions, but they could spontaneously generate prediction minimization. As a future study, I would explore modeling embodied beings based on the HGMs, to finally connect time scales with various cognitive phenomena including dreaming, synesthesia, and consciousness.

4.4 Conclusion

I first demonstrated that robust behavior, observed as vicarious trial-and-error (VTE), can be self-organized from hierarchical time scales, i.e., sensory processing, learning speed, and behavior (Chapter 2). I also presented that the neural network to generate VTE (i.e., conflict) showed activity patterns that worked on various time scales. This suggests the efficiency of having multiple time scales toward robust control of behavior. In Chapter 3, we showed that organization of multiple time scales is an emergent property of the biological neuron cultivated *in vitro*. To conclude, we demonstrated that, in both robotic systems and biological neurons, organization of multiple time scales is an emergent property, which will allow adaptive behavior.

Chapter 5

Acknowledgement

まず第一に、指導教官である池上高志教授に、心から感謝の意を述べさせていただきます。学部時代に池上教授の授業を拝聴した折、生命現象を物理的に理解する、複雑系物理の存在を知りました。それは自分の想像を遥かに越える世界の捉え方であり、学問を切り拓くこと、学者とはかくなるものかと深く感銘を受けました。大学院での５年間、本論文をまとめるにあたり、池上教授は越えるべき壁をいつも提示してくださいました。得られる結果は理想から遠く、私は初志を見失いがちでしたが、池上教授は常に、私に解くべき問題を示してくださいました。池上教授に教わったことは、これまでの人生の宝であり、私が今後科学者として研究を続ける動機です。池上教授にご指導、ご鞭撻を賜りましたこと、この上なくうれしく存じます。

池上研究室の皆様にも、心よりお礼を申し上げます。特に、先輩であり、共同研究者であるJulien Hubert 氏には、私が修士課程 1 年生の頃から、実験の組み立て方、論文の書き方、結果の読み解き方などを相談させていただきました。時に強情だった私に対し、どんなときも親切にアドバイスをくださり、コンピュータが壊れた時には、夜が遅くても、いつも直してくださいました。研究生活における日常的なことを話したり、将来の相談をしたり、よき友でもあったHubert 氏に大変助けられ、研究の日々を送りました。岡瑞起博士は、共同研究を通して、研究の進め方、論文の書き方を教えてくださり、また、研究費の申請、学会発表の資料作りに際しましてもご助言をくださいました。女性研究者としてご活躍されている岡博士は、私の憧れで、研究生活の過ごし方、将来の進路など、あらゆる面で相談に乗っていただき、研究者としての将来を具体的に思い描くきっかけを与えてくださいました。研究を進める中で、袋小路に迷いこむことが多くありましたが、こうして博士論文にまとめることが出来ましたのは、岡博士がいつも私を気にかけ、励ましてくださったおかげです。Nathaniel Virgo 博士は、論文執筆、研究内容に関して多くのご助言をくださいました。要を得ない

私の話に耳を傾け、整然としない私の論文を添削して、論文の筋道を一緒に考えてくれました。

高橋宏知講師には、深く感謝を申し上げます。第３章の実験において、培養神経細胞の挙動を解析する、大変貴重な機会を与えていただきました。実際の神経細胞は、コンピュータシミュレーションで行うものとは異なる挙動を示し、驚きと、知的な刺激に満ちたものでした。三田毅氏には、培養ニューロンのデータについて、また、解析について議論をさせていただき、大変お世話になりました。心から、お礼を申し上げます。

東京大学教養学部広域科学科の、８年来の同級生達にも、深くお礼を申し上げます。宇宙物理、生態学などそれぞれ異なる分野に進まれましたが、研究内容について、また研究生活に関して、いつもお力を貸してくださいました。学部時代に共に学び、議論をしたことは今も忘れません。志を共有する仲間を得られたこと、今後研究を続ける上で、大変心強く存じます。

本論文の審査をしてくださいました、國吉康夫教授、嶋田正和教授、開一夫教授には、心より感謝を申し上げます。

本研究は、Japan Society for the Promotion of Science (DC2), Denso Corp, Grant-in-Aid for Scientific Research (24300080 and 23680050), Grant-in-Aid for Scientific Research on Innovative Areas (24120704) の援助を受けております。

Chapter 6

Supporting Information – Development of the neural network

A reward-oriented behavior emerges through Hebbian learning. In this section, we explain the underlying mechanism of the emerged behavior. As described in the method section, we found several possible parameter sets by the genetic algorithm to maximise the fitness value. Although those different parameter sets might result in various strategies, here we describe the most possible mechanism to generate the reward seeking behavior of the robot, by following Bovet's description [16, 21]. The way of the description is almost same as what is written in his Ph.D. thesis [21], but we think it might be difficult to access it, so that we describe the mechanism here in this supporting information.

As is noted in the method section, the experiments took place in two phases: a training phase and a testing phase. In the training phase, the robot learned a simple correlation between the IR, vision and motor modules. In the testing phase, the robot was required to solve a reward seeking task in a T-shaped maze. Here we describe first the learning mechanism of the training phase, and afterward the testing phase.

6.1 Training phase

During the training phase, the robot explored the maze with the cue and reward deactivated and with the motor speed fixed to the two variables $v_2 = (1, 0)$ or $(0, 1)$, which induce left and right

turns respectively. When it touched a wall, the robot was returned to its initial position and the motor speed was switched to the other value. This process was repeated 20 times. The purpose of this phase was to give the robot a chance to learn the correlations between the IR, vision, and motor modules. Especially, the "IR – motor" and "vision – motor" path are important to solve the succeeding test phase; The "IR – motor" path is able to induce wall avoidance behaviour, and the "vision – motor" path gives the robot an important association between optical flow and turning speed. In this paper, optical flow means the difference between the current visual image and the image τ time units previously. The detailed descriptions are in the next parts.

6.1.1 "IR – motor" path

This path lets the robot avoid the walls. As is noted in the method section, the neuron has two components, i.e., the virtual and the state units. Activity patterns for each virtual and state units are illustrated in Fig. 6-1. In this figure, each module has five boxes, which corresponds to the five types of variables important for computing the neural network in Eq. 2 and 3 in the original manuscript, i.e.,

1. the state unit at the current time step $\boldsymbol{x}^m(t)$,
2. the state unit τ step before $\boldsymbol{x}^m(t-\tau)$,
3. the difference of those two $\Delta\boldsymbol{x}^m(t) = \boldsymbol{x}^m(t) - \boldsymbol{x}^m(t-\tau)$,
4. the virtual unit at the current time step $\tilde{\boldsymbol{x}}^m(t)$,
5. the difference between the virtual and state units $\Delta\tilde{\boldsymbol{x}}^m(t) = \tilde{\boldsymbol{x}}^m(t) - \boldsymbol{x}^m(t)$.

The IR module has circles in each of the five boxes, which denote activity levels of each of the six IR neurons. The left three circles corresponds to the left side of the IR sensors, while the right ones are for the right IR sensors. As for the motor module, the arrows inside the five boxes indicate the direction of turning. The virtual unit of the motor module $\tilde{\boldsymbol{x}}^{motor}(t)$ is used for calculation of the actual turning speed (Eq. 1).

When the robot turns toward a wall (Fig. 6-1A), to the left wall for example, this causes a correlation between an increase of the left part of the IR proximity signals and turning left.

Therefore the robot learns a positive correlation between $\Delta \boldsymbol{x}^{IR}(t)$ with the left part positive and motor activity $\boldsymbol{x}^{motor}(t)$ corresponding to turning to left (Eq. 2). Conversely, turning away from the left wall (Fig. 6-1B) causes a correlation between a decrease of IR signal in the left sensors and turning to right. Therefore, the robot will learn a positive correlation between turning left and the left part of the IR signal, and a negative one between turning right and the left part of the IR signal.

After learning this correlation, wall avoidance behavior is induced by the following process (Fig. 6-1C):

1. The robot makes a turn to the left wall, and gets close to it.
2. The left part of the IR signals increases.
3. Elements of the state unit $\boldsymbol{x}^{IR}(t)$ corresponding to the left IR sensors take positive values.
4. The virtual unit $\tilde{\boldsymbol{x}}^{IR}(t)$ is normally not activated, as there will be no learned correlation from the other modules to the IR module.
5. The difference between the virtual and state unit $\Delta \tilde{\boldsymbol{x}}^{IR}(t)$ $(= \tilde{\boldsymbol{x}}^{IR}(t) - \boldsymbol{x}^{IR}(t))$ has therefore its elements negative, which propagates through the positive connection from the IR module to the motor module learned above, and cause a turn to the opposite direction, i.e. the right turn.

Therefore, when the robot comes close to the wall, this "IR – motor" path causes a turn in the opposite direction, so that it can result in wall avoidance behavior.

6.1.2 "Vision – motor" path

This path lets the robot learn a correlation between optical flow and turning speed. In this paper, optical flow means the difference between the current visual image and the image τ time units previously (τ is the time delay in Eq. 2 and 3). The omni directional camera receives a grayscale image of the environment, so that it detects the black wall at the back and the white walls in the other sides of the T-maze. We also use the phrase "optical flow" to refer to the movement of the black wall image. For instance, if the image of the black wall moves in the right direction

in its visual field, then we call it optical flow in the right direction. By this “vision – motor” path, the robot can learn a correlation between optical flow and turning speed, i.e., turning left is correlated with optical flow in the right direction, and vice versa. This correlation is used in the testing phase.

6.2 Testing phase

After learning the basic correlation between the IR, vision and motor modules, the robot starts the testing phase, i.e., solving the T-maze. The reward is placed in the left or right arm of the T-maze and changes its position every five trials, but for simplicity, we only describe here the case with the reward on the left arm. As described in the method section, two paths are important for generating the reward seeking behavior. One is the “reward – vision – motor” path, and the other is the “tactile – motor” path. Those two paths can conflict with each other after the reward position changes, which will cause VTE behavior. In the following, we describe the detailed neural activity.

6.2.1 “Reward – vision – motor” path

This path leads the robot to go to the previous reward position. Therefore, if the reward position is fixed, the robot can get the reward. But it is not sufficient to reach the reward when the reward position switches. Fig. 6-2 illustrates the processes of two example trials. At the beginning of the trials (Fig. 6-2A), the robot is placed on the starting point at the center arm of the maze, where the omni directional camera detects the black wall at the back. Then, the robot goes straight, because of the constant forward velocity (Eq. 1). The whisker sensors start detecting the tactile cue on the left side (Fig. 6-2B). Hebbian learning captures a correlation between the increase of the tactile stimulation on the left side and sensory signals of the other modules (i.e., $\Delta \boldsymbol{x}^{tactile}(t)$ and $\boldsymbol{x}^{m}(t)$ in Eq. 2; m represents modules other than the tactile one). After a short time, the robot reaches the intersection point, and the whiskers break contact with the tactile cue. There, the whiskers detect the decrease of the tactile signal, which cancels out most of the previously learned correlations. At the first trial, the robot does not learn much correlations about the tactile cue and it makes no large turn, so that the robot once hits the back wall. But

a slight difference of the right and left motors, because of asymmetry from the training phase, leads the robot to one side of the T-maze. Therefore, it reaches the reward or punishment only by chance. We explain the case of the reward here, but the reverse situation, where the robot gets to the punishment side, happens in an analogous way.

Fig. 6-3A shows neural activity of the reward, vision and motor modules when the robot gets the reward at the first trial. The notation is the same as in Fig. 6-1. The five boxes inside the vision module denote activity patterns of its visual field for the state and virtual units (upper three boxes), and the optical flow for the state difference $\Delta \boldsymbol{x}^{vision}(t)$ and the virtual difference $\Delta \tilde{\boldsymbol{x}}^{vision}(t)$ (lower two boxes). As for the reward module, the sign inside the boxes indicates the activity level. As described in the method section, the virtual unit of the reward module $\tilde{\boldsymbol{x}}^{reward}(t)$ is always set to 1.0, which is illustrated by the black triangle.

If it reaches the reward, the robot gets an increase of the reward signal, and at the same time the camera detects the black wall on the right side of its visual field. There, the robot learns a positive correlation between $\Delta \boldsymbol{x}^{reward}(t)$ and $\boldsymbol{x}^{vision}(t)$ by Eq. 2 (Fig. 6-2C and 6-3A). This correlation stores the momentary camera image when the robot gets the reward, and the image gets remembered by the reward virtual unit $\tilde{\boldsymbol{x}}^{reward}(t)$, which is described next.

After getting the reward, the robot is returned to the starting position to start the second trial (Fig. 6-2D and 6-3B). By Eq. 3, the signal from the reward virtual unit first propagates through the synaptic connectivity to the vision module, as described in the previous paragraphs. Because the synaptic weights are positive, the signal from the reward virtual unit induces a similar activity pattern of the camera image at the reward point in the first trial, i.e., the black wall image on the right part of its visual field. Therefore, the virtual unit $\tilde{\boldsymbol{x}}^{vision}(t)$ has black on the right visual field. At the same time, the actual camera image, or the state unit $\boldsymbol{x}^{vision}(t)$, has its central visual field black, because the robot is facing the black wall. Therefore, the difference $\Delta \tilde{\boldsymbol{x}}^{vision}(t)$ (i.e., the difference between the virtual and state units) has a visual flow in the right direction. By Eq. 3, this optical flow in the virtual unit propagates into the motor module to induce a left turn, because of the correlation from the right optical flow to the left turn, which is learned in the training phase (Fig. 6-2D and 6-3B). Therefore, at the second trial, the robot is again successful. If the robot goes to the punishment at the first trial, a similar process makes the robot reach the reward side at the second trial.

So far, it becomes obvious that, by using the reward, vision and motor modules, the robot goes to the same side as the reward side during the first trial. But the robot cannot succeed when the position of the reward switches. To overcome this, the robot has to learn the correlation between the tactile cue and the reward. Next, we describe the learning with the whisker sensors.

6.2.2 "Tactile – motor" path

This path lets the robot follow the tactile cue which indicates the correct reward position. Here we explain the second trial, i.e., the trial after the robot learns the "reward – vision – motor" path, which leads it to turn to the previous reward position (Fig. 6-2D). Fig. 6-4A illustrates the learning process of this correlation. The sign in the boxes at the tactile module denotes reward(+), punishment(-) and nothing(0) for the left and right sides of the whiskers. In Fig. 6-4A, the robot makes a left turn because of the "reward – vision – motor" path described above and, at the same time, it breaks contact with the tactile cue. Therefore, by Eq. 2, the robot learns a correlation between the decrease of the state unit $\Delta \boldsymbol{x}^{tactile}(t)$, and the state unit $\boldsymbol{x}^{motor}(t)$ that induces a left turn.

After the second trial (Fig. 6-4B), if the robot touches the tactile cue on the left side, the state unit of the tactile module $\boldsymbol{x}^{tactile}(t)$ takes a positive value. The virtual unit of the tactile module $\tilde{\boldsymbol{x}}^{tactile}(t)$ is normally not activated because any correlations from the other modules to the tactile one are learned less often. Therefore, the difference of the virtual and state unit $\Delta \tilde{\boldsymbol{x}}^{tactile}(t)$ takes a negative value. By Eq. 3, this negative activity propagates through the learned connections to induce left turn.

These two neuronal paths have significant roles on solving the task. Additionally Bovet mentioned in his Ph.D. thesis [21] a connection that has a secondary role: the "tactile – vision" path can support the "reward – vision – motor" path. Any other connections are not mentioned in Bovet's paper, and they may not be necessary to solve the task. As described in the method and the result section, we cut those redundant connections, and proved that the robot was still able to solve the maze. However, we observed no VTE with this minimal connectivity, which let us to conclude the redundant connectivity is necessary for generating VTE.

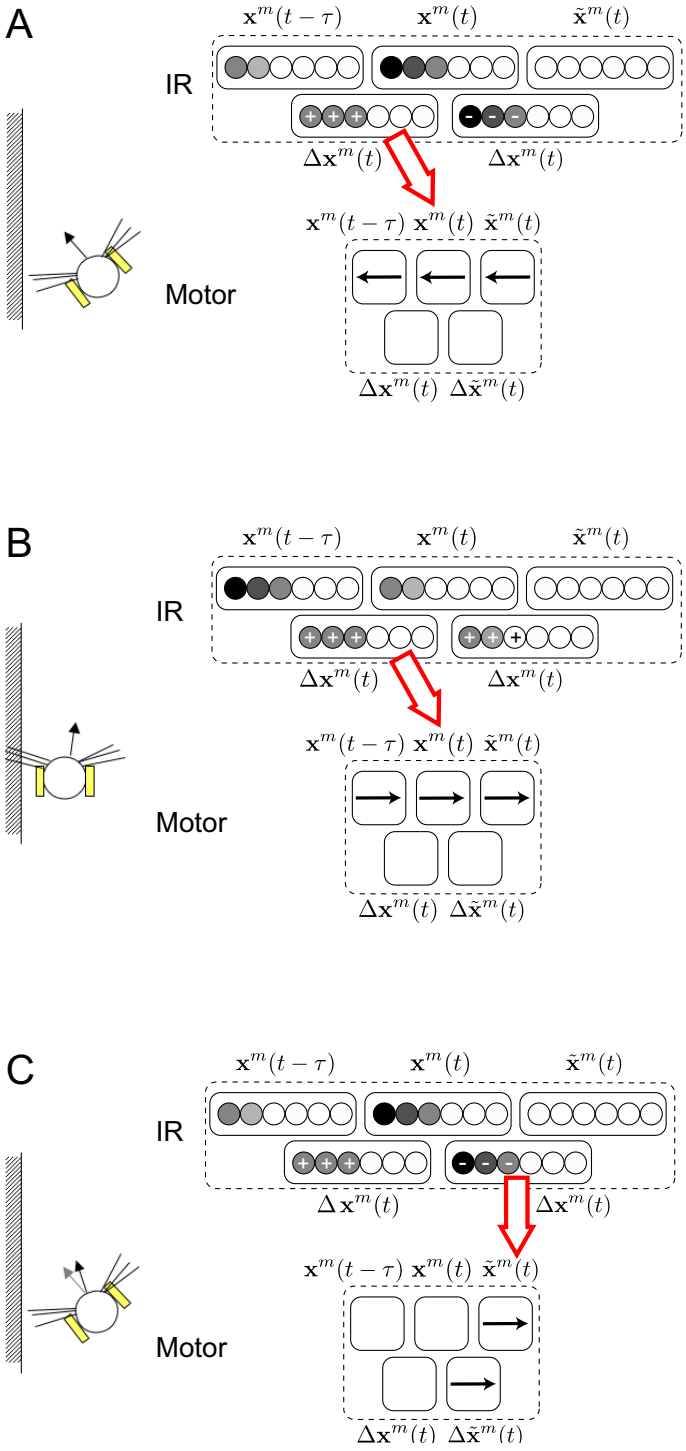

Figure 6-1: **Illustration of the learning process of the "IR – motor" path during the training phase.** This path lets the robot avoid the walls, and is learned in the training phase. Five boxes in each module show activities of each of the five types of variables used for computation of Eq. 2 and 3. The IR module has circles in each of the five boxes, which denote activity levels of each of the six IR neurons. In the motor module, the arrow indicates the direction of turning. A: When the robot turns to a wall, to the left wall for example, the robot learns a correlation between an increase of IR proximity signal in the left sensors and motor activity corresponding to turning to left. B: Conversely, it learns a correlation between a decrease of IR signal in the left sensors and motor activity corresponding to turning to right. C: After learning the correlation above, if it gets close to the left wall by making a left turn, the difference between the virtual and state units propagates into the virtual unit of the motor module, which generates an opposite turn (i.e., a right turn).

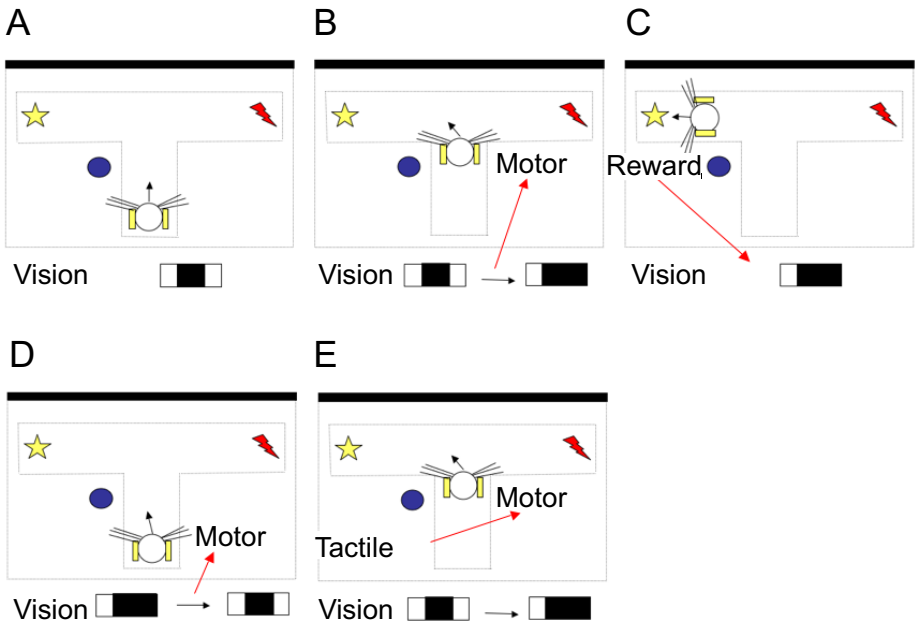

Figure 6-2: **The process of learning.** A, At the beginning of each trial, the robot is placed on the central arm of the T-maze. The camera detects the black wall at the back, and the central part of its visual field is activated. B, The robot has not get learned much correlations about the tactile cue so that the robot makes no large turn. But a slight difference in the right and left motors, because of asymmetry from the training phase, leads the robot to one side of the T-maze. C, When the robot reaches the reward, the Hebbian learning rule reinforces the correlation between an increase of reward and the camera image of the black wall. D, After getting the reward, the robot is returned to the starting position to start the second trial. Through the "reward – vision – motor" path learned at the first trial, the virtual unit of the reward module (always set to 1.0) propagates into the vision module, inducing a left turn, i.e., a turn to the previous reward position. E, During the turning, the left whiskers perceive the tactile cue, so that a "tactile – motor" correlation is produced.

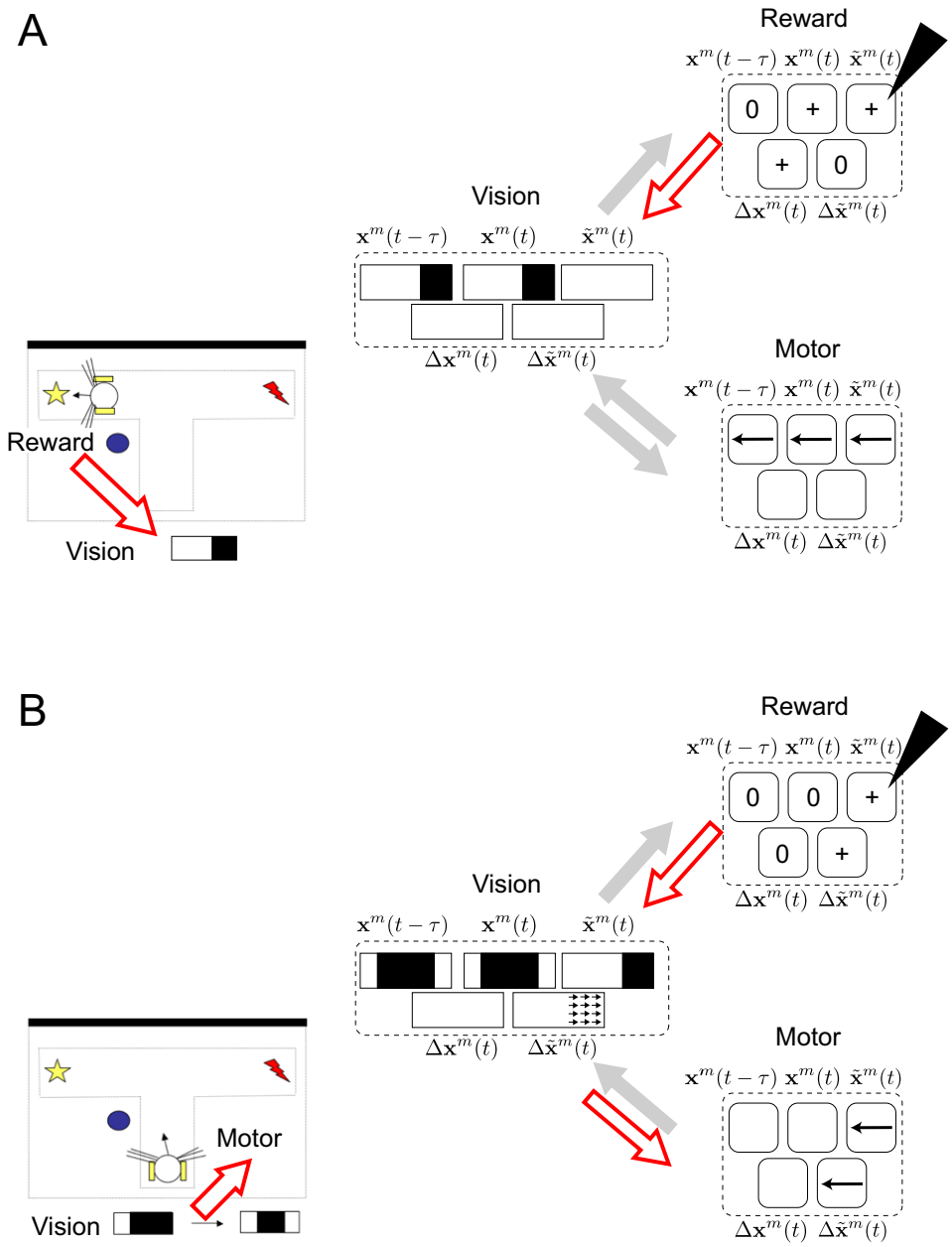

Figure 6-3: **Illustration of the learning process of the "reward – vision – motor" path during the first and second trials.** Five boxes in each module show activities of each of the five types of variables used for computation of Eq. 2 and 3. The five boxes inside the vision module denote activity patterns of its visual field (upper three boxes) or optical flow (lower two boxes). As for the reward module, the sign inside the boxes indicates its activity level. The virtual unit of the reward module $\tilde{\boldsymbol{x}}^{reward}(t)$ is always set to 1.0, which is illustrated by the black triangle. A, Neural activity when the robot gets the reward at the first trial. When it reaches the reward, the robot gets an increase of the reward signal (denoted by '+'), and at the same time the camera detects the black wall on the right side of its visual field. The robot learns a positive correlation between $\Delta\boldsymbol{x}^{reward}(t)$ and $\boldsymbol{x}^{vision}(t)$. B, Neural activity at the beginning of the second trial. By Eq. 3, the signal from the reward virtual unit $\boldsymbol{x}^{reward}(t)$ propagates through the synaptic connectivity to the vision module learned in A, to induce in the virtual unit of the vision module $\tilde{\boldsymbol{x}}^{vision}(t)$ a similar activity pattern of the camera image at the reward point in the first trial, i.e., the black wall image on the right part of its visual field. At the same time, the actual camera image (or the state unit $\boldsymbol{x}^{vision}(t)$) has its central visual field black. Therefore, the difference between them (or $\Delta\tilde{\boldsymbol{x}}^{vision}(t)$) has a visual flow in the right direction. This optical flow in the virtual unit propagates into the motor module to induce a left turn, because of the correlation learned in the training phase. Therefore, at the second trial, the robot is again successful.

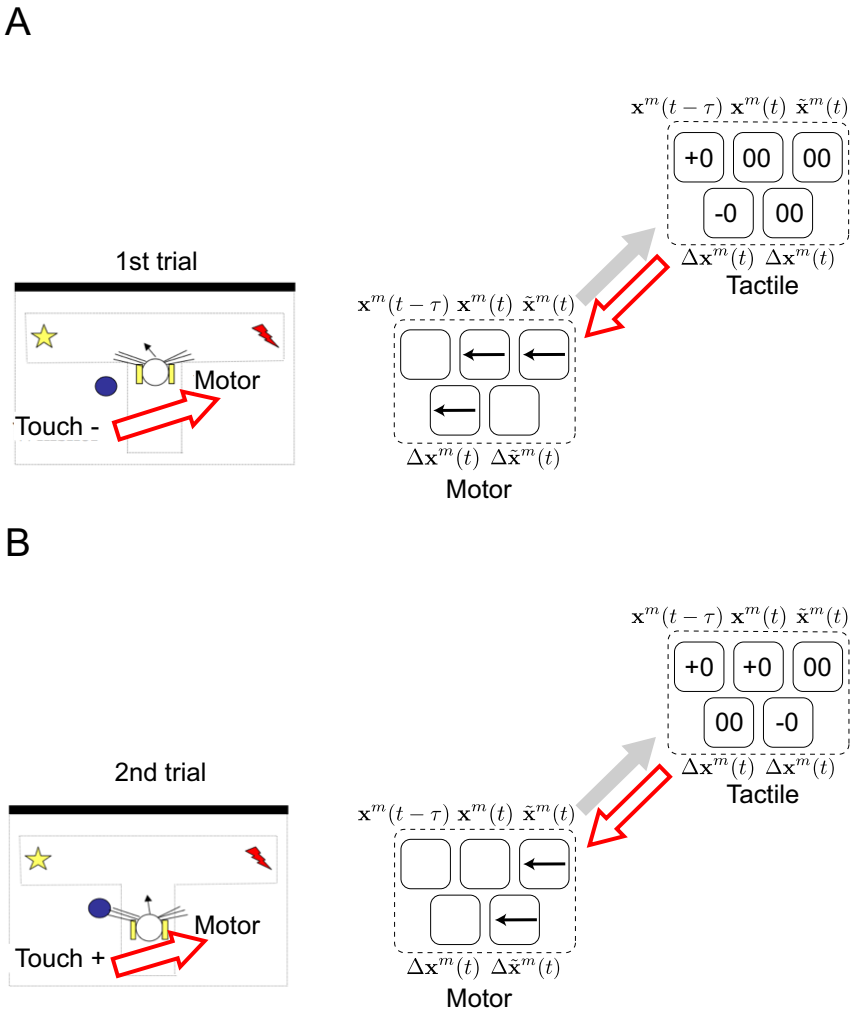

Figure 6-4: **Illustration of the learning process of the "tactile – motor" path in the second trial.** Five boxes in each module show activities of each of the five types of variables used for computation of Eq. 2 and 3. The sign in the boxes at the tactile module denotes reward(+), punishment(-) and nothing(0) for the left and right sides of the whiskers. A, Because of the "reward – vision – motor" path, the robot makes a left turn, and at the same time the left whiskers break contact with the tactile cue. Therefore, the robot learns a correlation between the decrease of the state unit $\Delta\boldsymbol{x}^{tactile}(t)$, and the state unit $\boldsymbol{x}^{motor}(t)$ that induces a left turn. B, After the second trial, if the robot touches the tactile cue on the left side, the state unit of the tactile module $\boldsymbol{x}^{tactile}(t)$ takes a positive value. The virtual unit of the tactile module $\tilde{\boldsymbol{x}}^{tactile}(t)$ is normally not activated because any correlations from the other modules to the tactile are not learned often. Therefore, the difference of the virtual and state unit $\Delta\tilde{\boldsymbol{x}}^{tactile}(t)$ takes a negative value. This negative activity propagates through the learned connections to induce a left turn.

Bibliography

[1] Berlekamp, E., Conway, J. H., and Guy, R. *Winning ways for your mathematical plays*. Academic Press, New York, 1982.

[2] Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. Autonomy: An information theoretic perspective. *Biosystems 91*, 2 (2008), 331–34.

[3] Boettiger, C. A., and Doupe, A. J. Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron 31*, 5 (2001), 809–818.

[4] Bovet, S. Emergence of insect navigation strategies from homogeneous sensorimotor coupling. In *Proceedings of the 9th International Conference on Intelligent Autonomous Systems* (2006), pp. 525–533.

[5] Bovet, S., and Pfeifer, R. Emergence of delayed reward learning from sensorimotor coordination. In *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)* (2005), Edmonton, pp. 841–846.

[6] Bovet, S. I. *Robots with Self-Developing Brains*. PhD thesis, Universität Zürich, 2007.

[7] Braitenberg, V. *Vehicles. Experiments in synthetic psychology*. MIT Press, 1984.

[8] Brooks, R. A. Intelligence without representation. *Artificial Intelligence: Building Embodied, Situated Agents 47*, 1–3 (1991), 139–159.

[9] Brooks, R. A. New approaches to robotics. *Science 253*, 5025 (1991), 1227–1232.

[10] Brooks, R. A. *Cambrian intelligence: The early history of the new AI*. A Bradford Book, 1999.

[11] Buhusi, C. V., and Meck, W. H. What makes us tick? functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience 6* (2005), 755–765.

[12] Buonomano, D. V. the biology of time across different scales. *Nature Chemical Biology 3*, 10 (2007), 594–597.

[13] Caporale, N., and Dan, Y. Spike timing-dependent plasticity: A hebbian learning rule. *Annual Review Neuroscience 31* (2008), 25–46.

[14] Carr, C. E. Processing of temporal information in the brain. *Annual Review of Neuroscience 16* (1993), 223–243.

[15] Cassenaer, S., and Laurent, G. Hebbian stdp in mushroom bodies facilitates the synchronous flow of olfactory information in locusts. *Nature 448* (2007), 709–713.

[16] Cheng, B., and Titterington, D. M. Neural networks: A review from a statistical perspective. *Statistical Science 9*, 1 (1994), 1–148.

[17] Clark, A. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36*, 03 (2013), 181–204.

[18] Collins, S., Ruina, A., Tedrake, R., and Wisse, M. Efficient bipedal robots based on passive-dynamic walkers. *Science 307*, 5712 (2005), 1082–1085.

[19] Crick, F. The recent excitement about neural networks. *Nature 337* (1989), 129–132.

[20] Czeisler, C. A., Duffy, J. F., Shanahan, T. L., Brown, E. N., Mitchell, J. F., Rimmer, D. W., Ronda, J. M., Silva, E. J., Allan, J. S., Emens, J. S., Dijk, D.-J., and Kronauer, R. E. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science 284*, 25 (1999), 2177–2181.

[21] Dautenhahn, K. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B 362*, 1480 (2007), 679–704.

[22] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B 39*, 1 (1977), 1–38.

[23] Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M. A. R., Kram, R., and Lehman, S. How animals move: An integrative view. *Science 288*, 5463 (2000), 100–106.

[24] Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., and Petersen, S. E. A dual-networks architecture of top-down control. *Trends in cognitive sciences 12*, 3 (2008), 99–105.

[25] Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., Burgund, E. D., Grimes, A. L., Schlaggar, B. L., and Petersen, S. E. A core system for the implementation of task sets. *Neuron 50* (2006), 799–812.

[26] Edelman, G. *Bright Air, Brilliant Fire: On The Matter Of The Mind*. Basic Books, 1992.

[27] Edwards, C. J., Alder, T. B., and Rose, G. J. Auditory midbrain neurons that count. *Nature Neuroscience 5*, 10 (2002), 934–936.

[28] Egger, V., Feldmeyer, D., and Sakmann, B. Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in rat barrel cortex. *Nature Neuroscience 2* (1999), 1098–1105.

[29] Eytan, D., and Marom, S. Dynamics and effective topology underlying synchronization in networks of cortical neurons. *The Journal of Neuroscience 26(33)* (2006), 8465–8476.

[30] Fair, D. A., equal contributor, A. L. C., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Schlaggar, B. L., and Petersen, S. E. Functional brain networks develop from a "local to distributed" organization. *PLOS COMPUTATIONAL BIOLOGY 5* (2009), e1000381.

[31] Feldman, D. E. Timing-based ltp and ltd at vertical inputs to layer ii/iii pyramidal cells in rat barrel cortex. *Neuron 27* (2000), 45–56.

[32] Feldman, H., and Friston, K. J. Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience 4*, 215 (2010), 1–23.

[33] Frey, U., Sedivý, J., Heer, F., Pedron, R., Ballini, M., Mueller, J., Bakkum, D. J., Hafizovic, S., Faraci, F. D., Greve, F., Kirstein, K.-U., and Hierlemann, A. Switch-matrix-based high-density microelectrode array in cmos technology. *IEEE Journal of Solid-State Circuits 45(2)* (2010), 467–482.

[34] Friston, K. Hierarchical models in the brain. *PLoS Computational Biology 4*, 11 (2008), e1000211:1–24.

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

[36] Friston, K. J., Trujillo-Barreto, N., and Daunizeau, J. Dem: A variational treatment of dynamic systems. *NeuroImage 41* (2008), 849–885.

[37] Froemke, R. C., and Dan, Y. Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature 416* (2002), 433–438.

[38] Fyfe, C., and Baddeley, R. Non-linear data structure extraction using simple hebbian networks. *Biological Cybernetics 72* (1995), 533–541.

[39] Gallistel, C. R. *The Organization of Learning*. MIT press, 1990.

[40] Gardner, M. Mathematical games: The fantastic combinations of john conway's new solitaire game "life". *Scientific American 223* (1970), 120–123.

[41] Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. A neuronal learning rule for sub-millisecond temporal coding. *Nature 383* (1996), 76–78.

[42] Gravner, J., and Griffeath, D. Cellular automaton growth on $\mathbb{Z}^2$: Theorems, examples, and problems. *ADVANCES IN APPLIED MATHEMATICS 21* (1998), 241–304.

[43] Hanczyc, M. M., Toyota, T., Ikegami, T., Packard, N., and Sugawara, T. Fatty acid chemistry at the oil-water interface: Self-propelled oil droplets. *J. Am. Chem. Soc. 129*, 30 (2007), 9386–9391.

[44] Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. Evolutionary robotics: the sussex approach. *Robotics and Autonomous Systems 20* (1997), 205–224.

[45] Hebb, D. O. *The Organization of Behavior*. Wiley, New York, 1949.

[46] Holland, J. H. *Adaptation in Natural and Artificial Systems*. Ann Arbor : The University of Michigan Press, 1975.

[47] Hu, D., and Amsel, A. A simple test of the vicarious trial-and-error hypothesis of hippocampal function. *Proceedings of the National Academy of Sciences of the United States of America 92* (1995), 5506–9.

[48] Iida, F., Pfeifer, R., Steels, L., and Kuniyoshi, Y. *Embodied Artificial Intelligence*. Springer, 2004.

[49] Iizuka, H., and di Paolo, E. Extended homeostatic adaptation: Improving the link between internal and behavioural stability. In *Proceedings of the 10th International Conference on Simulation of Adaptive Behavior* (Berlin, Germany, 2008), Springer-Verlag, pp. 1–11.

[50] Ikegami, T. Simulating active perception and mental imagery with embodied chaotic itinerancy. *Journal of Consciousness Studies 14* (2007), 111–125.

[51] Izhikevich, E. M. Simple model of spiking neurons. *IEEE Transactions on Neural Networks 14*, 6 (November 2003), 1569–1572.

[52] Johnson, A., and Redish, A. D. Neural ensembles in ca3 transiently encode paths forward of the animal at a decision point. *The Journal of neuroscience : the official journal of the Society for Neuroscience 27* (2007), 12176–89.

[53] Johnson, A., Varberg, Z., Benhardus, J., Maahs, A., and Schrater, P. The hippocampus and exploration: dynamically evolving behavior and neural representations. *Frontiers in human neuroscience 6* (2012), 216.

[54] Kacelnik, A., and Brunner, D. Timing and foraging: Gibbon's scalar expectancy theory and optimal patch exploitation. *Learning and Motivation 33*, 1 (2002), 177–195.

[55] Kohonen, T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics 43*, 1 (1982), 59–69.

[56] Kondo, S., Iwashita, M., and Yamaguchi, M. How animals get their skin patterns: fish pigment pattern as a live turing wave. *The International Journal of Developmental Biology 53* (2009), 851–856.

[57] Kuniyoshi, Y., and Sangawa, S. Early motor development from partially ordered neural-body dynamics: experiments with a cortico-spinal-musculo-skeletal model. *Biological Cybernetics 95*, 6 (2006), 589–605.

[58] Kuniyoshi, Y., and Suzuki, S. Dynamic emergence and adaptation of behavior through embodiment as coupled chaotic field. In *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2004), IEEE, pp. 2042–2049.

[59] Kuniyoshi, Y., Yorozu, Y., Ohmura, Y., Terada, K., Otani, T., Nagakubo, A., and Yamamoto, T. From humanoid embodiment to theory of mind. *Embodied Artificial Intelligence* (2004), 202–218.

[60] Lakoff, G., and Núñez, R. *Where mathematics comes from : how the embodied mind brings mathematics into being*. Basic Books, 2000.

[61] Langton, C. G. Self-reproduction in cellular automata. *Physica D 10*, 1–2 (1984), 135–144.

[62] Langton, C. G. Studying artificial life with cellular automata. *Physica D: Nonlinear Phenomena 22*, 1–3 (1986), 120–149.

[63] Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) – the unconscious initiation of a freely voluntary act. *Brain 106* (1983), 623–642.

[64] Lipson, H., and Pollack, J. B. Automatic design and manufacture of robotic lifeforms. *Nature 406* (2000), 974–978.

[65] Lizier, J. T., Heinzle, J., Horstmann, A., John-Dylan, Haynes, and Prokopenko, M. Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fmri connectivity. *Journal of Computational Neuroscience 30*, 1 (2011), 85–107.

[66] Lloyd, D. P. C. Post-tetanic potentiation of response in monosynaptic reflex pathways of the spinal cord. *The Journal of General Physiology 33*, 2 (1949), 147–170.

[67] Maeda, E., Robinson, H. P. C., and Kawana, A. The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *Journal of Neuroscience 15*, 10 (1995), 6834–6845.

[68] Mauk, M. D., and Buonomano, D. V. The neural basis of temporal processing. *Annual Review Neuroscience 27* (2004), 307–340.

[69] Mitchell, M. *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press, 1996.

[70] Moiseff, A., and Konishi, M. Neuronal and behavioral sensitivity to binaural time differences in the owl. *The Journal of Neuroscience 1*, 1 (1981), 40–48.

[71] Mondada, F., Bonani, M., Raemy, X., Pugh, J., Cianci, C., Klaptocz, A., Magnenat, S., Zufferey, J.-C., Floreano, D., and Martinoli, A. The e-puck, a robot designed for education in engineering. In *Proceedings of the 9th Conference on Autonomous Robot Systems and Competitions* (2009), pp. 59–65.

[72] Muenzinger, K. F. Vicarious trial and error at a point of choice: A general survey of its relation to learning efficiency. *Journal of Genetic Psychology 53* (1938), 75–86.

[73] Muenzinger, K. F., and Gentry, E. Tone discrimination in white rats. *J. Comp. Psychol. 12* (1931), 195–206.

[74] Nakajima, K., and Ikegami, T. Dynamical systems interpretation of reversal of subjective temporal order due to arm crossing. *Adaptive Behavior 16* (2008), 129–147.

[75] Nessler, B., Pfeiffer, M., Buesing, L., and Maass, W. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLoS Computational Biology 9*, 4 (2012), e1003037.

[76] Nolfi, S., and Floreano, D. *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT press, 2000.

[77] of Morphogenesis, U. C. B. Turing, a. m. *Philosophical Transactions of the Royal Society B: Biological Sciences 237*, 641 (1952), 37–64.

[78] Ogai, Y., and Ikegami, T. Microslip as a simulated artificial mind. *Adaptive Behavior 16* (2008), 129–147.

[79] Oka, M., and Ikegami, T. Exploring default mode and information flow on the web. *PLOS ONE 8*, 4 (2013), e60398.

[80] Packard, N. H., and Wolfram, S. Two-dimensional cellular automata. *Journal of Statistical Physics 38*, 5/6 (1985), 901–946.

[81] Pennartz, C. M. A. The ascending neuromodulatory systems in learning by reinforcement: comparing computational conjectures with experimental findings. *Brain Research Science 21* (1996), 219–245.

[82] Pennartz, C. M. A. Reinforcement learning by hebbian synapses with adaptive thresholds. *Neuroscience 81*, 2 (1997), 303–319.

[83] Pfeifer, R., and Bongard, J. C. *How the body shapes the way we think*. MIT press, 2006.

[84] Pfeifer, R., and Iida, F. Morphological computation: connecting body, brain, and environment. *Japanese Scientific Monthly 58*, 2 (2005), 48–54.

[85] Pfeifer, R., and Scheier, C. Sensory-moor coordination: The metaphor and beyond. *Robotics and Autonomous Systems 20* (1997), 157–178.

[86] Pfeifer, R., and Scheier, C. *Understanding intelligence*. M IT Press, 1999.

[87] Prigogine, I. *From Being to Becoming: Time and Complexity in the Physical Sciences*. W H Freeman & Co (Sd), 1981.

[88] Richelle, M., and Lejeune, H. *Time in Animal Behavior*. Pergamon, 1980.

[89] Rosenstein, M. T., Collins, J. J., and Luca, C. J. D. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D 65* (1993), 117–134.

[90] Rossler, O. E. Fraiberg–lenneberg speech. *Chaos, Solitons & Fractals 4*, 1 (1994), 125–131.

[91] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. *Parallel Distributed Processing* (1986).

[92] Rumelhart, D. E., and McClelland, J. L. *Parallel distributed processing: explorations in the microstructure of cognition*. MIT Press, 1986.

[93] Schirmer, A. Timing speech: a review of lesion and neuroimaging findings. *Cognitive Brain Research 21*, 2 (2004), 269–287.

[94] Schmidt, B., Papale, A., Redish, A. D., , and Markus, E. J. Conflict between place and response navigation strategies: Effects on vicarious trial and error (vte) behaviors. *Learning & Memory 20* (2013), 130–138.

[95] Schreiber, T. Measuring information transfer. *Phys Rev Lett 85* (2000), 464–464.

[96] Seth, A. K. A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synaesthesia. *Cognitive Neuroscience* (in press).

[97] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal 27*, 3 (1948), 379–423.

[98] Sims, K. Evolving virtual creatures. In *SIGGRAPH '94 Proceedings* (1994), pp. 15–22.

[99] Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron 32*, 6 (2001), 1149–1164.

[100] Sohn, M.-H., and Carlson, R. A. Implicit temporal tuning of working memory strategy during cognitive skill acquisition. *American Journal of Psychology 116*, 2 (2003), 239–56.

[101] Song, S., Miller, K. D., and Abbott, L. F. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience 3* (2000), 919–926.

[102] Soon, C. S., Brass, M., Heinze, H.-J., and Haynes, J.-D. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience 11*, 5 (2008), 543–545.

[103] Staniek, M., and Lehnertz, K. Symbolic transfer entropy. *Phys. Rev. Lett. 100*, 15 (2008), 158101.

[104] Sumbre, G., Muto, A., Baier, H., and ming Poo, M. Entrained rhythmic activities of neuronal ensembles as perceptual memory of time interval. *Nature 456*, 6 (2008), 102–107.

[105] Sun, J.-J., Kilb, W., and Luhmann, H. J. Self-organization of repetitive spike patterns in developing neuronal networks *in vitro*. *European Journal of Neuroscience 32* (2010), 1289–1299.

[106] Tarsitano, M. Route selection by a jumping spider (portia labiata) during the locomotory phase of a detour. *Animal Behaviour 72* (2006), 1437–1442.

[107] Thompson, A. Silicon evolution. In *Stanford University* (1996), MIT Press, pp. 444–452.

[108] Tolman, E. C. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev. 46* (1939), 318–336.

[109] Tzounopoulos, T., Kim, Y., Oertel, D., and Trussell, L. O. Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience 7*, 7 (2004), 719–725.

[110] Ulam, S. Statistical mechanics of cellular automata. *Proceedings of the International Congress on Mathematics 2* (1952), 264–275.

[111] van der Meer, M. A., Johnson, A., Schmitzer-Torbert, N. C., and Redish, A. D. Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron 67* (2010), 25–32.

[112] van Pelt, J., Wolters, P. S., Corner, M. A., and Rutten, W. L. C. Long-term characterization of firing dynamics of spontaneous bursts in cultured neural networks. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING 51(11)* (2004), 2051–2062.

[113] Varela, F., Thompson, E., and Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*. MIT press, 1991.

[114] Von Neumann, J., and Burks, A. W. *Theory of self-reproducing automata*. Illinois: University of Illinois Press Urbana„ 1966.

[115] Voss, J. L., Warren, D. E., Gonsalves, B. D., Federmeier, K. D., Tranel, D., and Cohen, N. J. Spontaneous revisitation during visual exploration as a link among strategic behavior, learning, and the hippocampus. *Proceedings of National Academy of Sciences 108*, 31 (2011), E402–E409.

[116] Wagenaar, D., DeMarse, T. B., and Potter, S. M. Meabench: A toolset for multi-electrode data acquisition and on-line analysis. In *Conference Proceedings of 2nd International IEEE EMBS Conference on Neural Engineering* (Pasadena, CA, 2005), pp. 518–521.

[117] Wolfram, S. Universality and complexity in cellular automata. *Physica D. 10* (1984), 1–35.