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

Eiko Matsuda[1], Julien Hubert[1] and Takashi Ikegami[1]

[1]The university of Tokyo, 3-8-1, Komaba, Meguro-ku, Tokyo, 153-8902, Japan
{eiko, jhubert, ikeg}@sacral.c.u-tokyo.ac.jp

Abstract

Vicarious trial-and-error(VTE) is a type of conflict-like behavior, observed in route selection tasks (Tolman (1939)). Studies of VTE have shown a correlation between the number of VTEs exhibited by a system with its learning efficiency. At the onset of learning a task, the number of VTEs increases, and when the learning reaches its plateau, it decreases.
The question we explore in this paper concerns the role of VTE. Basing ourselves on a model developed by Bovet and Pfeifer (2005), we ran robotic experiments to compute the number of VTEs during the learning of a T-maze task. Our results first show that what has been found in rats can be replicated in artificial systems. Furthermore, by changing the connectivity pattern of the original model, we discovered that the connection between VTEs and learning efficiency might not be necessarily true as our results show that two models exhibiting the same performance can possess a different pattern of VTEs. By comparing the robustness of the two models under varied conditions, we propose that VTEs are connected to the adaptivity of a system to environmental changes.

Introduction

In his experiments, Tolman (1939) observed that rats are seemingly hesitating when they must choose between one of two rooms, one of which containing a reward while the other being empty. The only cue differentiating the rooms is the color of their doors. A black door indicates the room provides a reward, and a white color indicates an empty room. To reach the reward, the rats must learn the relationship between the color of the door and the presence of the reward. During the learning phase, the rats have been seen moving their head from one door to another which is referred by Tolman as a conflict-like behavior named "vicarious trial-and-error (VTE)". In his experiments, Tolman noticed that the number of VTEs increases at the onset of the learning phase to start decreasing when the performance reaches its plateau. From that observation, VTE has been connected to learning efficiency.

Following Tolman's observations, other researchers started paying attention to the presence of VTEs in their studies. Hu and Amsel (1995) showed hippocampal contribution to VTEs. Johnson and Redish (2007) reported the presence of VTEs in experiments on rats who were shown to be simulating their next decisions internally before acting. Tarsitano (2006) found that, in a detour task, jumping spiders display two phases of action: the inspection phase, where spiders stop and inspect possible routes toward a target, and the locomotory phase, where spiders move toward a single direction. VTEs have been observed during the inspection phase. Tarsitano concluded that "one can speculate that it is a small but significant jump to use trial and error vicariously when choosing a goal to approach". Ikegami (2007) suggested the relationship between VTEs and private simulation. From these researches, VTE seems to have some essential role in internal reflection and decision making. However, the role of the VTEs has yet to be fully investigated.

The question we explore in this paper concerns the role of VTE. Using a model developed by Bovet and Pfeifer (2005) for T-maze learning experiments on robotic platforms, we study the presence of VTEs during the acquisition of the task. Our results display the same pattern of increase followed by a decrease in the number of VTEs as observed in the rat. Additionally we vary environmental parameters as well as the connectivity of the network in order to study the variations in the number of VTEs. Based on our results we hypothesize that VTEs might be connected to robustness and adaptivity. We first detail the environmental setup and the neural model in the next two sections. Then the results will be presented with a discussion of their significance.

Methodology

Our work is based on a robotic and neural model developed by Bovet and Pfeifer. The model combines five types of modalities to control a robot in a T-maze task. The neural model is self-organized with no hierarchy between modalities, nor predetermined sensori-motor relationship. Modalities are associated through Hebbian learning only (Hebb (1949)).

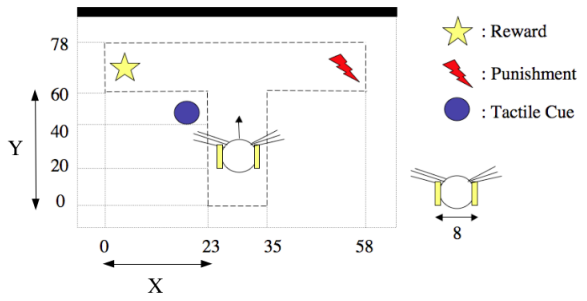

Figure 1: T-maze environment used for the experiment. At the beginning of each trial, the robot is placed on the central arm of the maze. The circle at the choice point represents the tactile cue, the star at one end of the maze indicates reward, and the lightning at the other end of the maze stands for punishment. The back wall is painted black and the other walls are white, which are detected by the robot's omnidirectional camera. Walls of the T-maze are perceived by the robot's proximity sensors. The length and the width of the T-maze are denoted by 'X' and 'Y'.

Experimental setup

The environment is a T-maze with one central arm and two side ones (see figure 1). A reward is located at the end of one arm, and a punishment is placed on the opposite one. The robot learns to reach the reward following a tactile cue placed at the end of the central arm, on the same side as the reward.

The robot is modeled following the e-puck robot (Mondada et al. (2009)) and is equipped with the following sensors and motors:

1) Tactile sensors: Tactile stimulation comes from 32 whiskers attached to the left and right sides of the robot. The signal is binary, on or off. Whisker sensors detect the tactile cue at the intersection point of the T-maze. The walls of the T-maze are low enough so that the whiskers can only detect the cue.

2) Vision sensors: Visual stimulation reflects the activity of the omnidirectional camera, which return grayscale values standardized from 0 to 1. This camera is composed of 20 pixels aligned horizontally. Everything in the T-maze is made white or transparent, except for the black back wall. In other words, the omnidirectional camera gets positive signals only from the black wall at the back of the T-maze. By this modality, the robot acquires destination information.

3) Proximity sensors: Six proximity sensors are regularly attached to the front half of the body. These sensors detect the distance from the robot to the walls of the T-maze. The values are standardized between 0 and 1. These sensors are attached low enough to only detect the walls of the T-maze. This modality is involved in the experiment only indirectly to achieve wall avoidance.

4) Reward sensitivity: The reward sensitivity is usually set to 0. It is raised to 1 to signal a reward and lowered to -1 to indicate punishment. The value is dependent on which side of the maze is reached by the robot.

5) Motors: The forward velocity of the robot v_f is constant and positive, and the turning degree v_t is determined by the neural controller of the robot which is explained later. Both v_t and v_f are standardized between 0 and 1, and activate actual left and right wheel velocities, v_l, v_r, as following:

$$\begin{pmatrix} v_l \\ v_r \end{pmatrix} = C \cdot \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} v_f \\ v_t \end{pmatrix} \quad (1)$$

where C is a constant for converting the standardized value to the actual motor speed. If $v_t > 0$, then $v_l < v_r$, which makes the robot turn left, and $v_t < 0$ produces a right turn.

As a training phase, the robot runs randomly in an empty maze with no tactile cue nor reward signals. Afterward, the tactile cue and the reward are introduced into the T-maze, and the robot must complete the task. The robot learns the correlation between modalities through Hebbian learning while acquiring a reward seeking behavior (for more detailed explanations, see Bovet and Pfeifer (2005)).

Neural Network - Bovet et al.'s Original Model

The neural network is composed of five modality modules: tactile, vision, proximity, reward, and motor (figure 2(a)). Each of them plays a separate role in treating the signals from its corresponding sensor (or motor) on the robot. Each modality has five types of neural populations, described in figure 2(b). These five populations are composed of the same number of artificial neurons, this number varying depending on the type of modality. For instance, the tactile modality has 32 neurons for each of the five populations while the motor modality has only one neuron per population. The five types of populations are described as follows:

1) Current state The current state of modality M, $\mathbf{x}^M(t) = (x_1^M(t), x_2^M(t), ..., x_m^M(t))$, receives signals from the corresponding sensors (or motors). For instance, tactile stimuli from 32 whisker sensors activate the corresponding 32 nodes of the current state.

2) Delayed state The delayed state $\breve{\mathbf{x}}^M(t) = \mathbf{x}^M(t-\tau)$ receives signals τ timestep in the past.

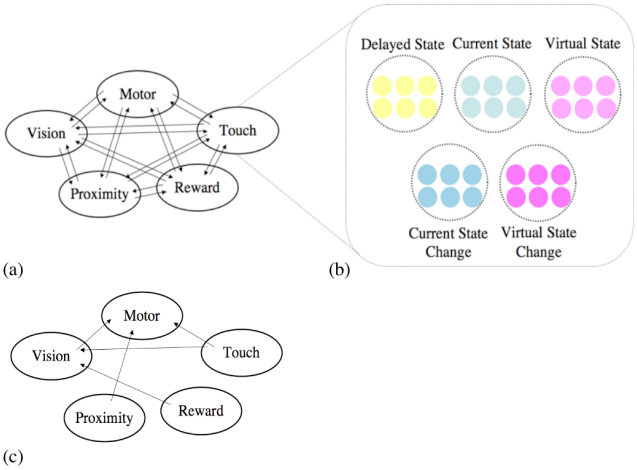

Figure 2: (a) Five sub-systems for each modalities make up the whole cognitive system of the robots. These modalities are fully connected with each other. (The original model) (b) Five types of neural populations, called the current state, the delayed state, the current state change, the virtual state, the virtual state change. (c) A new neural network where the five modalities are sparsely connected sparsely. (The minimal model)

3) Current state change The current state change $\mathbf{y}^M(t)$ is the difference between the current and the delayed state, as described below:

$$\begin{aligned}\mathbf{y}^M(t) &:= \mathbf{x}^M(t) - \breve{\mathbf{x}}^M(t) \\ &= \mathbf{x}^M(t) - \mathbf{x}^M(t-\tau)\end{aligned} \tag{2}$$

4) Virtual state The virtual state of modality M, $\tilde{\mathbf{x}}^M(t)$, is activated by the virtual state change of other modalities:

$$\tilde{\mathbf{x}}^M(t+1) := f(\Sigma_{N \neq M} W^{MN}(t) \cdot \tilde{\mathbf{y}}^N(t)) \tag{3}$$

where W^{MN} is the weight matrix connecting modality M to modality N and $f(x)$ is a sigmoid function, written as:

$$f(x) = \frac{1.0}{1.0 + exp(-a \cdot x)} \tag{4}$$

5) Virtual state change The virtual state change of modality M, $\tilde{\mathbf{y}}^M(t)$, is the difference between the virtual and the current state.

$$\tilde{\mathbf{y}}^M(t) := \tilde{\mathbf{x}}^M(t) - \mathbf{x}^M(t) \tag{5}$$

The current state population, delayed state population and current state change population do not possess any in or out connections toward other modalities. All virtual state change populations are connected to virtual state populations of the other modalities. For instance, the neurons from the virtual state change population of the tactile modality are connected to the neurons of the virtual state population of all the other modalities. Those connections are the only ones present in the model.

All the connections of the model are tuned using a modified version of Hebbian learning. The main difference between Hebb's version is that the pre and post synaptic neurons are not used to compute the change of the synaptic connections. For the learning, the neurons of the non-virtual populations are used. Instead of the virtual state population, the neurons of the current state population are used for the Hebbian learning. Similarly, the neurons of the current state change population replace the ones from the virtual state change population. Mathematically, this corresponds to the following equations:

$$\begin{aligned}\Delta W^{MN}(t) &:= l \cdot (\mathbf{x}^N(t)\mathbf{y}^M(t)^T - \alpha \mid \mathbf{y}^M(t) \mid W^{MN}(t)) \\ W^{MN}(t+1) &= W^{MN}(t) + \Delta W^{MN}(t)\end{aligned} \tag{6}$$

where l is the learning rate and α is the forgetting rate. Because of the α, the weight between a pair of neurons is decreased if the two neurons are not activated at the same time. It also prevents the weights from growing to infinity.

To clarify the inner algorithm of the neural model, we detail the steps leading to the generation of the outputs as follows:

1. Sensory information is transferred to the current state population.
2. The delayed state population is updated, followed by the current state change population.
3. Hebbian learning is applied on all the connections of the model.
4. The activity of the neurons from the virtual state change populations of all modalities are propagated to the neurons of the virtual state populations using a feedforward algorithm (see equation 3).
5. The activation of the single neuron of the virtual state population from the motor modality is assigned to the output v_t from equation 1.

For additional details on this model, please refer to the original paper Bovet and Pfeifer (2005).

Neural network - Minimal model

In addition to the original neural network invented by Bovet and Pfeifer, we conducted experiments with a new neural network model. In the original model, modalities are fully connected (figure 2(a)), while our new model has only minimal connectivity among modalities. By "minimal" connectivity, we mean connections which have a specific role to

solve the task mentioned in Bovet's paper, as shown in figure 2(c). We expect that the behavioral difference between the original and the minimal connectivity model will allow us to uncover the role of VTE.

Setup of the Genetic Algoritm

Bovet and Pfeifer's model relies on the following parameters: learning rates and forgetting rates for each modalities, update frequency of the neural network, τ for the delayed state and constants for the sigmoid function in equation 4. Despite the authors not mentioning how to select those parameters, we found out that slight differences in their value can strongly influence the performance of the robot. This is partly due to our experiments adopting more tolerant conditions than the original experiment, like a broader T-maze. To tune these parameters and optimize the performance of the controller, we employ a genetic algorithm (GA)(Holland (1975)). Our GA possesses a population of 100 individuals to optimize 59 parameters using tournament selection, single point crossover applied with a probability of 70% and a 1% mutation rate. We also use elitism by simply copying the 5 best individuals directly to the next generation. A fitness function $F(t)$ at generation t is calculated as:

$$F(t) = \begin{cases} +5\ points,\ if\ it\ reaches\ the\ reward. \\ +0.25\ points,\ if\ it\ reaches\ the\ punishment. \\ +0\ points,\ if\ it\ gets\ timeout. \end{cases} \tag{7}$$

The amount of points assigned is determined arbitrarily. The trials are repeated 100 times from one fixed initial position, which gives a maximum fitness value of 500. We conducted several runs of the GA for the original and the minimal model respectively.

Results

For each model, we evolved 5 runs of GA. Figure 3 shows 2 out of 5 GA runs get the maximum fitness value (100% success) with the original model, and, in the case of the minimal model, 3 out of 5 GA runs successfully evolved. We selected one individual for each model from these evolved runs and counted the number of VTE they displayed.

Our methodology to count the number of VTE in a robot is similar to the one used by Tolman. In our case, the robot does not possess a head moving independently from its body so the whole body movement has to be considered. One VTE is granted if the turning degree v_t from the equation 1 changes its sign. In order to filter small oscillations around a turning degree of 0, a VTE is only granted if the sign change is outside the range $[-0.3; 0.3]$.

Figure 4 shows the number of VTEs observed for one evolved individual for each model. We can see from figure 4(a) that the robot evolved with the original connectivity model exhibits more VTEs at the beginning of the learning,

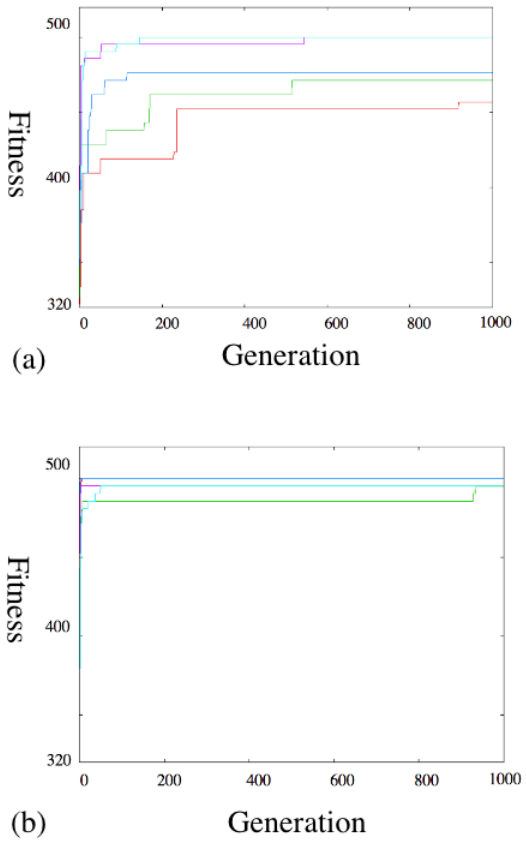

Figure 3: Fitness values of the five runs of GA. The X axis represents the number of generations, and the Y axis the fitness of the best individual in each generation. Maximum fitness value is 500. (a) In the case of the original model. Two out of five runs of GA get the maximum value. (b) In the case of the minimal model. Three out of five runs get the maximum value.

to decrease afterward. This observation is similar to Tolman's experiments on real rats (Tolman (1939); Muenzinger and Fletcher (1934)). On the other hand, the robot with the minimal connectivity model shows VTEs in a lower amount while remaining constant during the course of the experiment (figure 4(b)). Despite this difference in the number of VTEs, both models show a success rate of 100%. This result implies that VTEs are not directly related to performance in learning, but might have another purpose.

In order to study if the presence of VTEs could imply a higher level of robustness for the robot, we analyze the performance under varying initial conditions. During evolution, the starting position is $(x, y) = (29, 20)$. This experiment explores if the performance of the robot is affected by a change in its initial position by testing it from every other

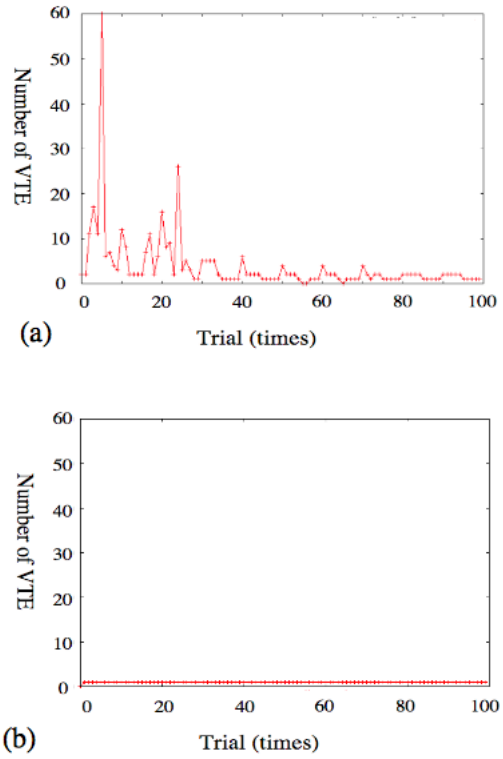

Figure 4: Change in the number of VTE during learning. (a) In the case of the original model. (b) In the case of the minimal model.

starting position inside the central arm of the T-maze. Each position has been tested 100 times to obtain the final results.

Figure 5 shows the results for each model. The first observation from this figure is that the performance is not constant over all starting positions. Some areas lead to higher performances. The second observation concerns the comparison of the variance of the performance between the two models. In the case of the original model, the variance remains under 400 while the minimum variance of the minimal model is around 600 as seen in figure 6. This means that, despite the two models having a similar average performance, the original model seems to withstand changes in starting position. On the other hand, the minimal model is strongly affected by the initial position. This result implies that the presence of VTEs could be associated with a higher level of robustness to changes in the environment.

Based on the success rate, we observed 5 different types of behaviors:

Going to the reward As we described above, the robot successfully reaches the reward side. In the case of the original model, the number of VTEs becomes higher at the beginning of the learning, and decrease afterward similarly to experiments with real rats (figure 4). But with the minimal model, we only observed lower and stable VTEs.

Going to the punishment With about 0 % success rates, the robot learns to reach the punishment side. As the learning progresses, the number of VTE increases and afterward decreases with the original model. In the case of the minimal model, we did not observe this VTE change.

Going to the same direction The robot learns to go to the same fixed side (right side or left side) and gets around 50 % success rates. The number of VTE remains high during the whole experiment.

Behavioral transition The robot transit among the three previous behaviors - going to the reward, the punishment, and the one side - and displays success rates between 30 % and 70%. This transition might have some relationship with chaotic itinerancy. However, this behavioral transition cannot be seen in the minimal model. The number of VTE remains high during the whole experiment.

Random The robot acts seemingly randomly and the success rate is around 50 %. The number of VTE remains high during the whole experiment.

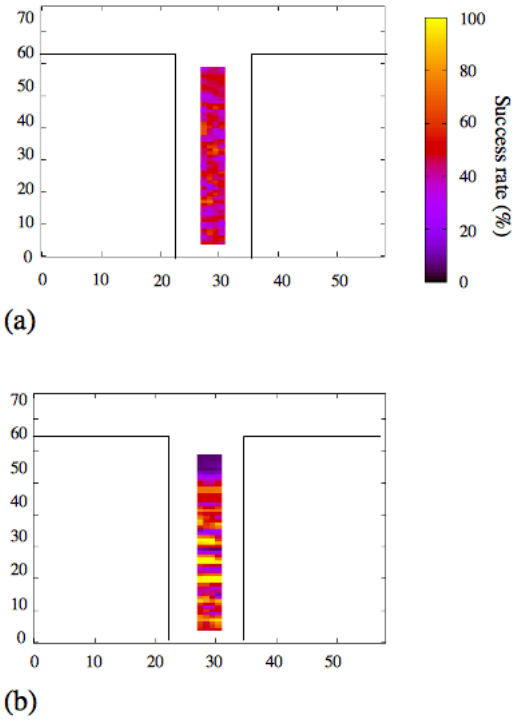

Figure 5: Average success rates for each starting positions. (a) In the case of the original model. (b) In the case of the minimal model.

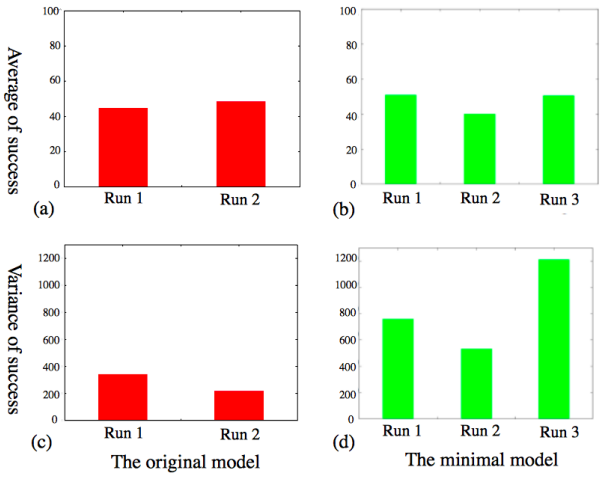

Figure 6: Average and variance of success rates per starting position. These graphs show the results of two (or three for the minimal model) GA runs of the original and the minimal model respectively. The red graphs present the results for the original model while the minimal model are in green. (a), (b) Average of success rates. (c), (d) Variance of success rates.

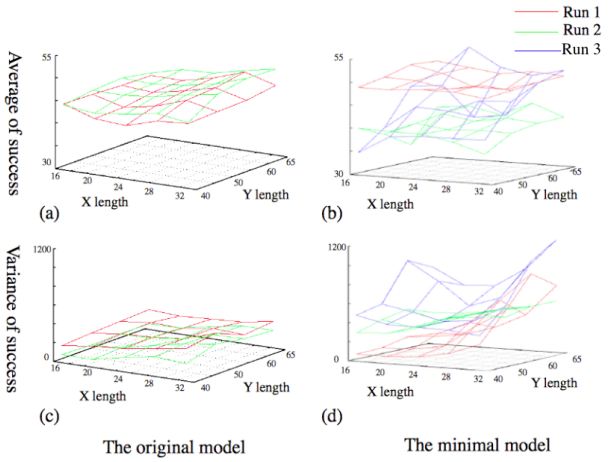

Figure 7: Average and variance of success rates per starting position, for different sizes of the T-maze. The size of the T-maze indicated by X and Y length corresponds to that of figure 1. Red and green lines corresponds to each two runs of GA. (a), (b) Average of success rates, for the original and the minimal model respectively. (c), (d) Variance of success rates.

In order to investigate further the robustness of the evolved controllers against environmental change, we carried out the same experiments with different T-maze sizes. We varied the width and the length of the T-maze, as drawn in figure 1, and calculated the average and the variance of the success rates for every initial positions. With the original model, the robot does not change its performance in respect to the average and the variance of the success rates. The robot with the minimal model gets affected by a slight change in environmental size (figure 7). This result confirms that the presence of VTE can be an indicator of the robustness of the neural system.

Conclusion

Our experiments aimed at uncovering the roles of VTEs through robotic experiments. Our work relies on a model developed by Bovet and Pfeifer where a neural network equipped with Hebbian learning commands a robot to complete a T-maze task using multiple sensory modalities (Bovet and Pfeifer (2005)). Unlike Bovet's work, we composed the whole setup in a computer simulation, and conducted experiments under varied conditions, while optimizing the parameters of the model using a GA. This setup allows us to compute the number of VTEs during the learning of the route selection task.

We compared two models, one with full connectivity among modalities, and the other with minimal connectivity. Although both models exhibit the same performance, or 100% success, the former shows similar VTE curves to experiments with real rats, while the latter does not. This implies that VTE might not be related to performance in learning but would rather be caused by a redundant connectivity pattern.

We also noticed the original model, or the model with redundant connectivity, maintains its success rate to about 50% in most cases, regardless of perturbations to initial conditions or environmental size, which is accompanied by more VTEs. On the other hand, the model with minimal connectivity exhibits a lower robustness against perturbations. This model shows almost no VTEs. In conclusion, we offer the hypothesis that VTE might be linked to adaptivity to environmental changes.

In addition, we observed three seemingly stable behavioral patterns, and behavioral transition among those three patterns. This transition might have something to do with chaotic itinerancy (Ikegami (2007)). However, the dynamics of the neural network has not been studied and the cause of the VTEs has yet to be uncovered. Additional studies of this model, such as analyses based on chaos theory (Ogai and Ikegami (2008); Nakajima and Ikegami (2008)), or analyses from the field of differential topology (Thom (1972)), could shed some lights on the mechanisms of VTEs.

Acknowledgements

We would like to thank Simon Bovet for his help on reproducing the original experiments. Julien Hubert thanks the Monbukagakusho Scholarship from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

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

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley, New York.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor : The University of Michigan Press.

Hu, D. and Amsel, A. (1995). 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*, 92:5506–9.

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

Johnson, A. and Redish, A. D. (2007). 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*, 27:12176–89.

Mondada, F., Bonani, M., Raemy, X., Pugh, J., Cianci, C., Klaptocz, A., Magnenat, S., Zufferey, J.-C., Floreano, D., and Martinoli, A. (2009). The e-puck, a robot designed for education in engineering. In *Proceedings of the 9th Conference on Autonomous Robot Systems and Competitions*, volume 1, pages 59–65.

Muenzinger, K. F. and Fletcher, F. (1934). Motivation in learning: I. electric shock for correct response in the visual discrimination habit. *J. Comp. Psychol.*, 17:266–277.

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

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

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

Thom, R. (1972). *Structural Stability and Morphogenesis*. W. A. Benjam.

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