Creation of a Robot that is Conscious of Its Experiences

S. Torigoe*. R. Igarashi*. T. Komatsu*.
J. Takeno*

**Meiji University, School of Science and Technology, Computer Science, Kanagawa, Japan (e-mail:juntakeno@gmail.com).*

Abstract: We have developed a robot that is capable of consciousness and emotions similar to humans. We represented consciousness in this robot, added emotions to assist in the evolution of the conscious system of the robot, and successfully performed mirror image cognition tests with the robot. Emotion is a basic function for the robot to enable it to avoid life-threatening situations. We now wish the robot to evolve and develop by itself. Humans continue to develop themselves by accumulating experiences. We believe that robots of the future should also be able to evolve and develop by themselves by repeating learning or accumulating experiences.

This paper reports on a robot that has functions similar to the consciousness and emotions of humans and is capable of avoiding physical danger to itself by remembering its past experiences.

Keywords: neural networks, recognition, robotics, consciousness, emotion, feelings, experiences

1. INTRODUCTION

The most mysterious subjects of humans are their brains and consciousness. These mysteries have been challenged by many researchers including psychologists, philosophers and brain scientists. As robot scientists we have also been studying these themes from various perspectives. Considering the remarkable progress in technological development in recent years, it would not be surprising if robots were able to communicate naturally with humans in the near future. Many researchers in robot science are studying consciousness. A number of research institutions and businesses are also developing conscious robots. The WE-4R II robot from Waseda University, for example, reportedly has reacted like having consciousness and emotions. The robot we are developing differs in that it represents its emotions using an artificial nerve model for the conscious system termed the MoNAD, for Module of Nerves for Advanced Dynamics. Many learning robots have been reported but our robot is also unique in that it learns its own experiences using emotions.

In the course of our study, we have established a definition for consciousness: consistency between cognition and behavior generates consciousness. Based on this definition, we have constructed a conscious system using recurrent neural networks. A robot with this conscious system successfully underwent imitation behavior and mirror image cognition was tested. This paper introduces a robot that is capable of avoiding obstacles using its experience-conscious function as it engages in typical imitation behavior. The authors demonstrate that by developing a robot with this function a new model of the human brain can be constructed and a robot with functions very similar to human consciousness will be developed in the near future.

Consciousness will be discussed first, followed by emotions. We will then discuss MoNAD, the conscious module we have devised. We will further describe the construction of the MoNAD-based conscious system and the imitation behavior capable of a robot using this system. The discussion will include references to Neural Network-based design and learning of experiences triggered by conscious emotions. The experiments we conducted with the robot and our observations of the results are reported in this paper.

2. ON CONSCIOUSNESS

Consciousness generally refers to a state where one knows what oneself and another are doing. Consciousness occurs when one is paying attention to something, and is thinking or awake. However, no complete definition of consciousness is available yet.

Our objective is to generate human-like consciousness in a robot. We believe that the following four cases are closely related to the consciousness of humans:

Case 1: Mirror neurons

Case 2: Mimesis theory

Case 3: Imitation behavior, a medical case

Case 4: Study of Imitation Behavior of Neonates by Meltzoff, Moore et al.

Cognition of the self is necessary to generate self-awareness. We therefore decided first to create the function of imitation and then to verify whether cognition of the self could be achieved from the state of the self and the state of the other obtained as feedback. We successfully developed the

MoNAD conscious system module which is capable of generating the imitation function and distinguishing between the self and others. We also developed the conscious system consisting of the MoNADs arranged in a hierarchy.

Using this conscious system, we symbolized the state of the self and the state of the imitating other, and compared them. If the two symbols were close to each other, the imitating other would be a conscious existence close to the self.

3. ON EMOTIONS

We propose the following hypothesis about emotions: Emotions occur upon receiving information from both inside and outside of one's physical body. We note that which is unpleasant to oneself and provide support to avoid behaviors that make others unpleasant. By offering this support, one can eliminate harmful obstacles by oneself. This concept was built into the robot that we developed.

4. STRUCTURE OF MoNAD

MoNAD, the conscious module that we developed, is a computation model for consciousness which uses neural networks.

The MoNAD (Fig. 1) consists of (a) cognition system, (b) behavior system, (c) primitive representation, (d) symbolic representation and (e) input/output (S and M).

External information S via the input passes through the MoNAD basically in the order of p1, (a), p5 and (d). At the symbolic representation (d), the information is used to represent the state of the self and the other using the language labels RL and BL. Information S can be an item of information coming from a lower-level module in the hierarchy. RL is the language label for information currently cognized, and BL the language label for the next behavior to be performed. RL is determined by neuro-computation of the information S, cognition information immediately before BL, and information on the behavior performed immediately before M'. In the absence of information from a higher-level module, the information of the cognition representation RL is just copied to the behavior representation BL. At the symbolic representation, the state of the self and the state of the other are language-labeled separately. Since our current research focuses on studying the function of human consciousness, we use supervised learning, or a top-down design, for the neural network.

Behavior commands from BL are transmitted through the MoNAD basically in the order of p6, (b), p2 and output M. M is an item of output information from the MoNAD that is generally given to the lower-level modules. The value for M is determined by neuro-computation of the information passing through BL and p6, input information S, and information on the behavior performed immediately before M'.

One of MoNAD's features is that the cognition system and behavior system share the same primitive representation. This sharing of the primitive representation makes it possible to learn behaviors when cognizing and learn cognition when behaving. The use of a closed information loop, in which the information circulates through the primitive representation and symbolic representation, assists in the generation of inner thought and expectations in the MoNAD. Another feature of the MoNAD is that somatic M' makes it possible for the system to cognize the state of the self.

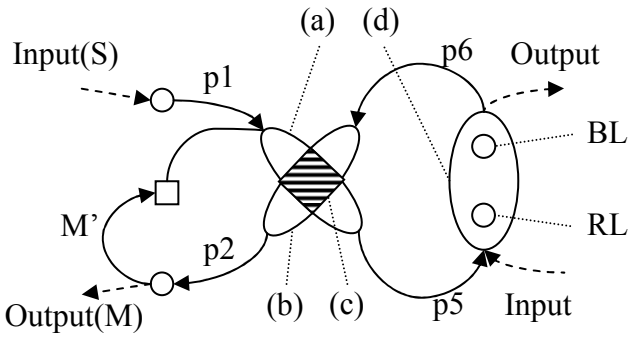

Fig. 1 MoNAD Structure

5. CONFIGURATION OF THE CONSCIOUS SYSTEM

The conscious system consists of three subsystems: reason subsystem, emotion-feelings subsystem, and association subsystem. Each subsystem has the MoNAD structure.

And all three subsystems have hierarchical structures with many MoNADs. The reason subsystem cognizes the external environment and the state of the self using the input information. It outputs a behavior command via the output M. The emotion-feelings subsystems represent emotions and feelings using the information of the state of the body. The top layer of the emotion-feelings subsystem hierarchy consists of two feeling MoNADs respectively representing Pleasant and Unpleasant feelings. Information from the reason subsystem (cognized language label) is also taken into consideration to correctly cognize the Pleasant and Unpleasant feelings.

Lastly, the association subsystem modifies the representation of both the reason and emotion and feelings subsystems using the information contained in these subsystems. The system learns through this process, and outputs behavior commands from the reason system that aims to eventually create a pleasant state for the self.

Figure 2 is Reason's MoNAD. Figure 3 is Emotion-Feeling's MoNAD. Figure 4 is Association's MoNAD.