A STUDY OF A CONSCIOUS ROBOT
An Attempt to Perceive the Unknown

Soichiro Akimoto and Junichi Takeno
Robot Science Laboratory, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa, 214-8571, Japan

Keywords: Cognition and Behavior, Consciousness, Pain of the heart, Detecting unknown, Neural Networks, Robotics.

Abstract: The authors are developing a robot that has consciousness, emotions and feelings like humans. As we make progress in this study, we look forward to deepening our understanding of human consciousness and feelings. So far, we have succeeded in representing consciousness in a robot, evolved this conscious system by adding the functions of emotions and feelings, and successfully performed mirror image cognition experiments using the robot. Emotions and feelings in a robot are, like those of humans, basic functions that can enable a robot to avoid life-threatening situations. We believe that consistency of cognition and behavior generates consciousness in a robot. If we can detect what happens in the robot when this consistency is lost, we may be able to develop a robot that is capable of discriminating between what it has learned and what it has not learned. Furthermore, anticipate that the robot may eventually be able to feel a "pain of the heart." This paper reports on autonomous detection by a robot of non-experienced phenomena, or awareness of the unknown, using the function of consciousness embedded in the robot. If the robot is capable of detecting unknown phenomena, it may be able to continually accumulate experiences by itself.

1 INTRODUCTION

We have already defined our belief that consciousness is generated from consistency of cognition and behavior. Based on this definition, we have developed a conscious system using recurrent neural networks (Suzuki, 2005). We then incorporated this conscious system into a robot, and the robot performed successfully in experiments on imitation behavior and mirror image cognition, or so-called mirror tests (Igarashi, 2007) ,(Takeno, 2008).

The present paper reports on our development of a robot that is capable of detecting a situation that it has never experienced before and representing a "heartfelt discomfort" or anxiety such as unpleasantness (a feeling) or pain (an emotion) according to the stimuli occurring in such a situation.

By developing a robot having these functions and creating a new model of the human brain based on the robot, the authors would like to demonstrate that it is possible to provide medical information on treatment of human brain diseases and develop a robot having a function very close to that of human consciousness in the near future.

2 PAIN OF THE HEART

This chapter first introduces phantom pain. A patient who has had a limb amputated due to injury or disease may feel as though the limb still existed as it was before amputation and may feel pain in the nonexistent limb. (Ramachandran, 1999)

The authors believe that the patient feels pain at the invisible and nonexistent limb because the brain cannot cognize the actual situation correctly, or what is cognized by the brain is different from the real world. The authors hypothesized that this cognitive gap generates unpleasant, and leads to a feeling of pain. If this hypothesis is true, we might then say that phantom pain—a feeling of unpleasant—occurs when the consistency of cognition and behavior (the definition of consciousness used by the authors used throughout our study) is lost.

The authors define the unpleasant and pain that occur when consistency of cognition and behavior is lost as "pain of the heart" for the purpose of elucidating phantom pain. We call this "pain of the heart" because it is the working of the reason subsystem whereas bodily pain is a result of physical stimuli, such as when the body collides with some external object.

Akimoto S. and Takeno J. (2010).
A STUDY OF A CONSCIOUS ROBOT - An Attempt to Perceive the Unknown.
In *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence - Artificial Intelligence*, pages 498-501
DOI: 10.5220/0002704904980501

Unpleasant arising from perception of the unknown, as discussed in this paper, is one of the "pain of the heart."

3 STRUCTURE OF A MONAD

A MoNAD is a conscious module developed by the authors. It is a computational model for studying consciousness using neural networks (NNs).

A MoNAD consists of a cognition system (a), behavior system (b), primitive representation region (c), symbolic representation region (d) which is a common area shared by the cognition and behavior systems, and input/output units (S and M) as shown in Fig. 1.

External information (S) enters the MoNAD at the input unit and normally goes through p1, cognition system (a), p5 and symbolic representation region (d). The symbolic representation region (d) represents the state of the self and the other as determined by the input information and language labels RL and BL. RL is a language label for information that is currently being cognized while BL is the label for the behavior that is to be performed next. RL is neuro-calculated from information S, cognition information BL one-step prior and behavior information M' one-step prior. Information of the cognition representation RL is constantly being copied to the behavior representation BL in the absence of any other information input from the higher-level modules into BL. In the symbolic representation region, the state of the self and the other is labeled separately. At present, all neural networks used in the study are top-down networks, and supervised learning is used because the main topic of this present study is the function of human consciousness.

The behavior command from BL basically passes through p6, behavior system (b), p2 and output unit M. M is the information output from the MoNAD. It is also transmitted to the lower-level MoNAD modules. Information M is neuro-calculated from information passing through BL and p6, input information S and behavior information M' one-step prior.

One of the features of our MoNAD is that the cognition system and behavior system share the primitive representation region (c). Thanks to this primitive representation region, the system learns behaviors when cognizing and learns cognition when behaving. The closed loop information circuit passing through the primitive representation region and symbolic representation region makes it possible for the MoNAD to generate artificial inner thoughts and expectations in the robot. Another feature of the MoNAD is that the state of the self is understood by somatic sensation M'.

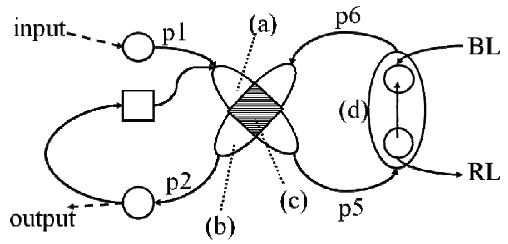

Figure 1: Structure of MoNAD.

4 STRUCTURE OF THE CONSCIOUS SYSTEM

The conscious system comprises three subsystems: reason, emotion-feelings and association. All subsystems have a similar MoNAD structure. The reason and emotion-feelings subsystems have a hierarchical structure formed of MoNADs. The reason subsystem uses input information to cognize the outer world and the state of the self, and outputs behavior from the output unit. The emotion-feelings subsystem employs information from the condition of the robot's body and represents emotions and feelings. Two feeling MoNADs, one representing pleasant and one unpleasant, form the top layer of the emotion subsystem. Information from the reason subsystem (cognized representation) is also used in the final determination of a state of pleasant or unpleasant.

The association subsystem uses information from the reason and emotion-feelings subsystems and integrates the representations of both subsystems. The association subsystem learns, by using the result of integration, and outputs a behavior capable of eventually realizing pleasant for the self from the reason subsystem.

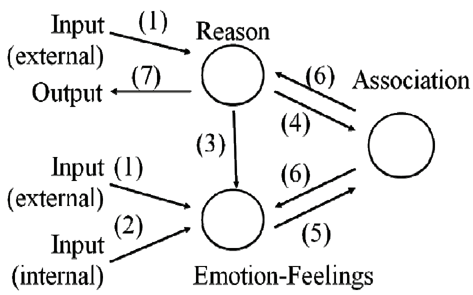

Figure 2: A Model of the Conscious System.

5 PERCEPTION OF THE UNKNOWN

To reproduce the situation where consistency of cognition and behavior is lost, or when there is a gap between the cognition and behavior, the authors conducted color identification experiments with the robot. The robot learns colors beforehand and discriminates between learned and non-learned colors. Two situations may be expected of the robot: one is where the pre-learned colors are cognized by the MoNADs smoothly and the other is where the non-learned colors are not cognized smoothly.

The robot pre-learns green, red and blue colors.

In the experiment, the robot is shown black, which is a non-learned information. We selected black because the robot would easily identify it to be unknown since black has no color information at all. We intend to try other complex colors later.

Images are taken by the camera embedded in the robot. The color is analyzed using the RGB values of the image. One image is taken about every second. This image-taking continues until the robot identifies the color. The robot represents pleasant from its feeling subsystem when it succeeds in identifying the color. Unpleasant is represented from the feeling subsystem when the robot fails to identify the color.

The feelings of the robot are classified into 4 types: much unpleasant, unpleasant, pleasant, much pleasant. Pleasant is determined when the robot identifies the color while unpleasant is determined when it fails.

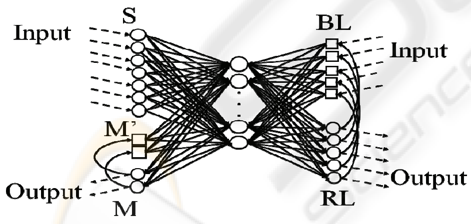

Figure 3: Development of the Conscious System with a MoNAD Structured Neural Network.

6 DETECTION OF GAPS BETWEEN COGNITION AND BEHAVIOR (DIFFERENCES BETWEEN RL AND BL)

To detect the gaps between cognition and behavior, we calculated the gaps in bit patterns between RL and BL (to be discussed later) in the primitive representation region. We expected that a gap between cognition and behavior would be present in the primitive representation region when consistency between cognition and behavior was lost because information on both cognition and behavior (inputs of external conditions, somatic sensations, and the result of the behavior one-step prior that is used to determine behavior) coexists in this region.

Calculating the bit pattern gap (or simply, the P-error) involves finding the sum of the mean squared error of the bits of RL and BL (note that each RL and BL have a 5-bit pattern). The first 3 bits represent the identified color and the last 2 bits are the behavior of the self.

We determine that no gap exists between cognition and behavior when the P-error is less than a certain value (0.0002).

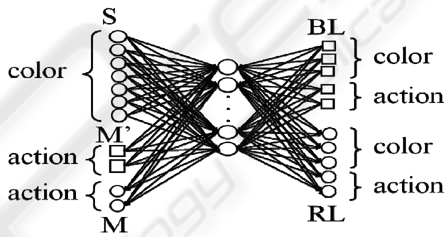

Figure 4: Description of RL and BL Nodes.

7 ROBOT EXPERIMENTS

An *e-puck* commercial mini-robot was used in our experiments. We installed color LEDs on the standard robot to be able to visualize its internal state.

The experiment was conducted in the following manner. First, colors are presented to the robot which is equipped with a camera. The robot views the color and behaves as it has learned. The robot then responds in one of four ways: advances (upon seeing green), stops (red), backs up (blue), or oscillates back and forth (non-learned color).

First, we placed a color in front of the camera of the robot to have the robot identify the color and respond according to what it has learned. We checked the feeling (pleasant or unpleasant) and representation of the robot as it identified the color successfully. Another color was then shown to the robot to continue the experiment.

The robot advances when green is shown. Green is a learned color and the robot cognizes it relatively quickly. When successful, the robot feels pleasant and the relevant LEDs light up.

The robot stops upon seeing red. Red is also a learned color like green and the robot cognizes it relatively quickly. When successful, the robot feels pleasant and the relevant LEDs light up.

The result is the same when blue is shown to the robot, and the robot backs up and feels pleasant.

When a non-learned color (black in this case) is shown, the robot oscillates back and forth. The robot performs many more computations than when viewing a learned color, as it attempts to cognize the unknown color. The robot eventually represents unpleasant in its emotion subsystem as the number of computations increases while it tries to identify the unknown color. The 'pain' LED lights up on the robot as it feels very unpleasant. The robot identifies the color.

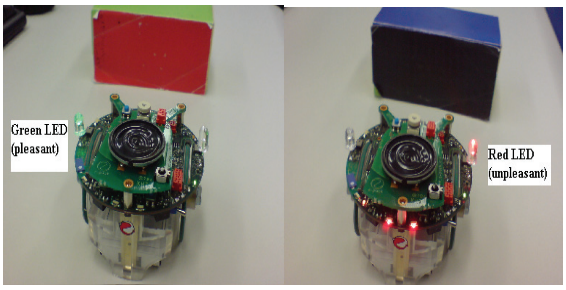

Figure 5: Views of Experiments.

8 RESULTS OF EXPERIMENTS AND OBSERVATIONS

For the three learned colors—green, red and blue—the consistency of cognition and behavior was established in the robot after 3 or 4 recursive computations with the MoNAD, and the colors were then successfully identified. The robot showed a pleasant state when successfully identifying colors. When presented with non-learned information (black), the robot required 5 to 10 recursive computations to identify the information, and the robot showed a very unpleasant state.

Figure 6 shows the P-error between RL and BL when the color was changed from red to green, and from blue to black, respectively, as well as from green, blue and black to red.

Thanks to our embedded conscious system used in these experiments, the robot showed, by representing unpleasant, that it was capable of perceiving non-learned information or a condition that it had never experienced before.

This was our demonstration of perceiving the unknown using a robot.

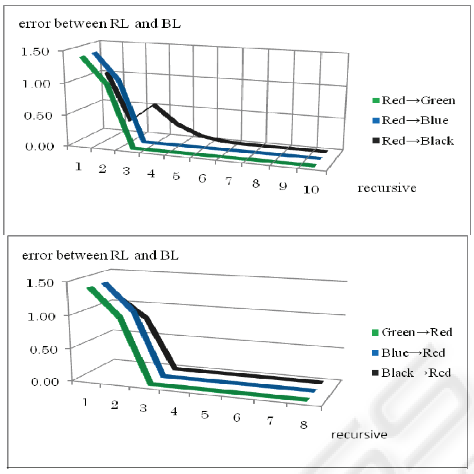

Figure 6: RL, BL and P-error.

9 SUMMARY

This paper discussed consciousness, emotions and feelings. Based on a belief that consistency between cognition and behavior generates consciousness, the authors believe that pain and unpleasant, which occur when consistency between cognition and behavior is lost, comprise "pain of the heart."

REFERENCES

V, S, Ramachandran., Sandra, Blakeslee., 1999. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. HARPER PERENNIAL, pp.3-328.

Suzuki, T., Inaba, K., Takeno, J., 2005. Conscious Robot that Distinguishes between Self and Others and implements Imitation Behavior. IEA/AIE, pp.101-110.

Igarashi, R, Suzuki, T, Shirakura, Y, Takeno, J., 2007. Realization of an emotional robot and imitation behavior. The 13th IASTEND International Conference on Robotics and Applications, RA2007, pp.135-140.

Takeno, J., 2008. A Robot Succeeds in 100% Mirror Image Cognition. INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS, VOL. 1, NO. 4, DECEMBER 2008, pp.891-911.

Torigoe, S, Igarashi, R, Komatsu, T, Takeno, J., 2009. Creation of a Robot that is Conscious of Its Experiences. ICONS2009 September 21-23, 2009 Istanbul, Turkey.