Procedia Computer Science
Volume 71, 2015, Pages 44–49
2015 Annual International Conference on Biologically Inspired Cognitive Architectures

Pleasant and Unpleasant States in a Robot

Haruki Ebisawa[1], Ryuma Matsushita[1] and Junichi Takeno[1*]
[1] *Robot Science Laboratory, Meiji University,*
1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8571, Japan.
h.ebisawa0224@gmail.com, ryumtst@gmail.com, juntakeno@gmail.com

Abstract
We have been conducting research with the objective of enabling a robot to perform human-like autonomous behavior and communication. For this purpose, we have developed consciousness modules, termed MoNADs, which by means of groups of neurons comprise the elements that make up the functions of consciousness, emotions and feelings that are possessed by humans, and mounted them on a robot. In this study, we propose a consciousness model that enables a robot to evolve in accordance with rules based on our definitions of pleasant and unpleasant states in a conscious system comprised of MoNADs. For the definitions of the pleasant and unpleasant states, we are using the "smoothness of the information flow" in the conscious system, where "pleasant" is the state in which the information that has come in to the robot is flowing smoothly, while the state in which information is not flowing smoothly is defined as "unpleasant." By means of these definitions, the robot becomes able to autonomously evolve using uniform and consistent rules in both of the systems of the brain, the sub-system that governs reason and the sub-system that is responsible for emotions and feelings.

Keywords: Emotion, Feelings, Conscious system, MoNAD, Pleasant, Unpleasant, Autonomously evolve

1 Introduction

Robots of today are required to have autonomous behavior, communication skills and other human-like qualities. Considering that robots possessing increasingly advanced intelligence and precision will be developed through future advances in technology, robots will appear that will exist and be familiar to humans, acting like partners or members of the family. And this requires them to have human-like qualities. We believe that human consciousness is the origin of human thought, feelings and behavior, and that we can achieve a robot that can think like humans by providing the robot with consciousness.

Up to now, human consciousness has been studied in such disciplines as philosophy, psychology, brain science and neuroscience. In addition, studies have been underway in a visible form of human consciousness through modeling that mimics artificial consciousness, in studies that give minds and

Selection and peer-review under responsibility of the Scientific Programme Committee of BICA 2015

doi:10.1016/j.procs.2015.12.192

personalities to machines (Holland, 2007), as well as in the cognitive architecture of self-conscious robots (Chella, 2009). From among those research studies, we have focused on mirror neurons (Gallese, 1996) and mimesis theory (Donald, 1991), and we have defined the origin of consciousness to be the "consistency between cognition and behavior." Based on our definition, we have devised a consciousness module called Module of Nerves for Advanced Dynamics (MoNAD) (Takeno, 2011) which is comprised of recursive neural networks (RNN). In our previous studies, a robot equipped with these MoNADs successfully completed self-recognition experiments, that is, mirror image cognition tests (Takeno, 2005). Moreover, by incorporating emotions and learning mechanisms in the system, we enabled the robot to successfully complete experiments in which it autonomously performed danger avoidance using its episodic memory (Komatsu, 2011). To be able to achieve a high degree of human-like consciousness and intelligence and enable the robot to evolve autonomously by interacting with its environment, it is preferable to define developmental rules that are unified throughout the entire system without distinguishing between the system involved with reason and the system involved with emotions and feelings. We thought about what was the foundation upon which humans evolved from birth and whether that could be the rule at the core of our system. This was our definitions of the "pleasant" and "unpleasant" states.

2 Definitions of Pleasant and Unpleasant States

Humans are always continually learning things in their daily lives. However, there are still many mysteries about learning behavior, and it can not be said that research is getting closer to any clarification. Up to now, there has been much progress in the fields of machine learning and evolutional learning using the techniques of artificial intelligence, and these have also been applied to the learning of robots. Further developments in artificial intelligence technology have been achieved through these studies. However, in observations of the learning behavior of humans, the relationships with emotions and feelings have been strongly pointed out. Although algorithms have also been proposed in the past to represent the feelings of a robot, such as from synthetic language and voice intonation generated by humans (Oudeyer, 2003), there has been little consideration of this topic. First of all, with respect to emotions and feelings, we took two states, "pleasant" and "unpleasant," as the basis, and we focused on the continual learning that humans do to change their own state from an unpleasant one to a pleasant one. This idea is very similar to the reward used in machine learning (Williams, 1987). However, in the context of the concept of a reward as mentioned here, it is not necessarily possible to point out any direct association to a benefit or non-benefit to the system itself.

Therefore, we decided to devise the new definitions of "pleasant" and "unpleasant" states in order to directly associate the concept of emotions and feelings to the system's own benefit and non-benefit. When the information that has entered the system is flowing smoothly through the system, this is defined as a "pleasant" state, and when the information is not flowing smoothly (including when the information is stagnant), this is defined as an "unpleasant" state. In other words, the "pleasant" and "unpleasant" states are not simply output values in the neural network, they represent the system as a whole. In this way, for example, by defining a state in which "unknown" information has entered the system as "unpleasant," and the state with "known" information as "pleasant," it is possible to enable the robot to evolve its knowledge with consistent rules. In other words, the robot feels that unknown information is unpleasant, and thus behaviorally learns so that it will know that information. And when the same information comes in to the system again, it is then determined to be known, which elicits a pleasant feeling.

To enable the robot to evolve its knowledge, in the future we will advance the discussion on known information and unknown information (Kushiro, 2011) as an issue to be dealt with while using pleasant and unpleasant feelings.

3 MoNAD Consciousness Module

The definitions of the “pleasant” and “unpleasant” states proposed in this paper can be achieved by the MoNAD consciousness module that we devised. The MoNAD is comprised of a number of parts: (a) input, (b) primitive representation, (c) cognitive representation (RL), (d) behavioral representation (BL), (e) output, and (f) somatic sensation unit (SSU). Three kinds of information are processed in the MoNAD: external information from the input, output cycled information entering the SSU, and behavior-changing information from higher layer MoNADs entering the behavioral representation (BL), and these are integrated in the primitive representation and the values of the cognitive representation (RL) and output are determined simultaneously. In other words, the MoNAD has achieved a structure in which, by means of primitive representation, it is simultaneously learning to perform a behavior when it cognizes while learning to cognize when it performs a behavior. In addition, depending on the primitive representation, RL and BL, artificial inner thought or expectation can be achieved with the information cycle paths. Also, the MoNAD has a feature in which it is conscious of its own state by means of somatic sensation information cycles.

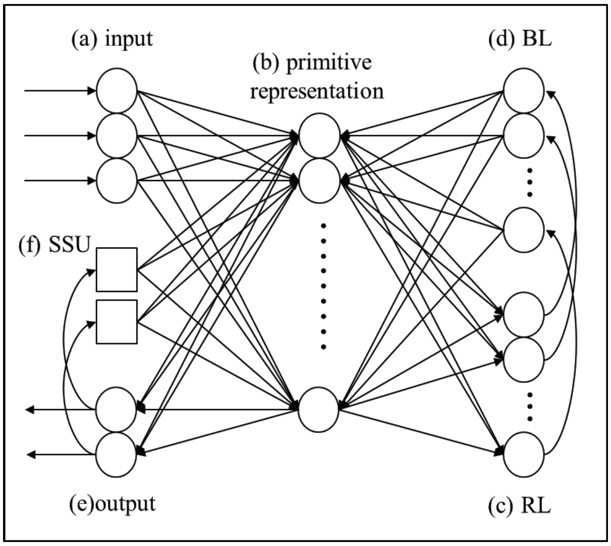

Figure 1: MoNAD consciousness module

4 MoNAD Characteristics

In order to enable a determination to be made as to whether the information coming into the robot is known information or unknown information, in this study we considered the process of information flow though information cycles.

The MoNAD has the following characteristics.

When the information that comes in to the system is continually and iteratively cycled between the cognitive representation and behavioral representation, the values of the cognitive representation and behavioral representation will ultimately converge and become equivalent. And when the information is known (i.e., the information has already been learned), compared to when the information is unknown (i.e., unlearned information), large differences can be observed in the condition of this iteration. When the information is known, the difference in values will tend to shrink all at once and the cycled information will flow, although the values will eventually settle at equivalence. But when the information is unknown, the difference between the values may not readily shrink and may oscillate. In addition, the difference in the values of the known and unknown information will further diverge due to the size of the value at each individual input node, making it possible to clearly distinguish between them. With these characteristics, it is possible for the MoNAD itself to determine whether information is known or unknown, and based on that information, to be conscious of pleasant and unpleasant emotions. As the representation of complex emotions is difficult still, we though of having these emotions represent the conscious system as a whole, and this is described in the following paragraphs.

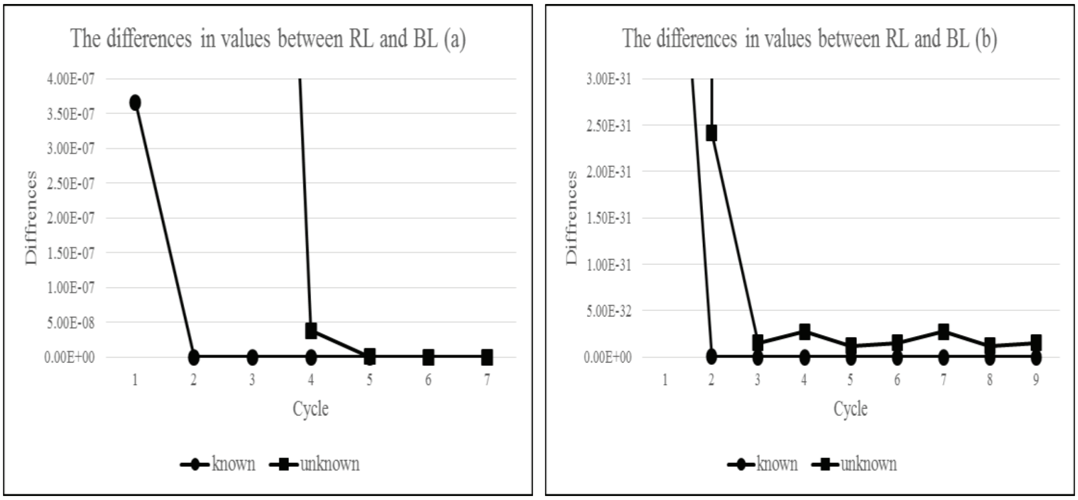

Table 1: Distinguishing between known and unknown information in a MoNAD
(a) Delay caused by unknown information
(b) Vibration caused by unknown information

5 Conscious System Determination of Pleasant and Unpleasant States

The conscious system consists of three sub-systems: reason system, emotions-and-feelings system, and association system, each of which is composed of a plurality of MoNADs. The reason system and emotions-and-feelings system are constructed in hierarchies using the respective MoNADs. Based on the input information, the reason system determines its own state in the external environment, and the emotions-and-feelings system represents emotions and feelings. The respective representation of those two types of information, from the reason system and from the emotions-and-feelings system, can be changed in the association system.

This conscious system consisting of MoNADs makes use of the characteristics of each of the MoNADs as described in the previous section. That is, if the information flowing through lower layer MoNADs (located close to the external environment) is deteremined to be known information, it can be easily transmitted to the linked upper layer MoNADs. However, if the information is determined to be unknown, it is not transmitted to the upper layer MoNADs, and that information becomes stagnant. In this manner, when the information is flowing to the upper layers of the conscious system, the state when the information is flowing smoothly is defined as “pleasant,” and the state when information is not flowing smoothly is defined as “unpleasant.” And if the input information is determined to be unknown information, the conscious system learns that information to make it known information. In other words, the conscious system self-evolves on its own so that “the cognitive and behavioral representations of MoNADs with stagnant information will be made more stable quickly.”

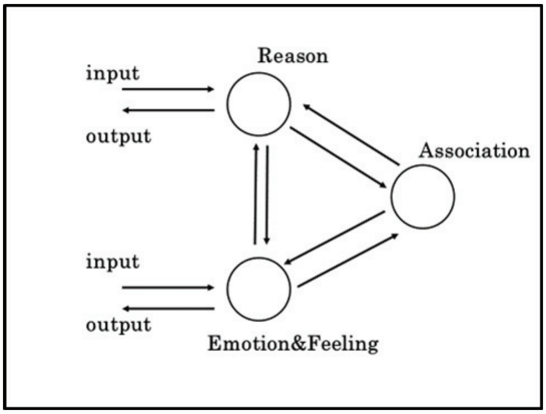

Figure 2: Conscious system model diagram

6 Discussion

The developmental rule of this conscious system is that it always learns unknown information to make it known information.

Why does learning unknown information to make it known information lead to the evolution of the system?

Human reactions differ with regard to things that are predictable as opposed to things that are not. The response to something predictable is derived instantaneously, and triggers a behavior. When something is not predictable, humans scrutinize it carefully and cautiously. This behavior is thought to arise from humans’ concern about their own safety. When it comes to their safety, the largest issues for humans are “life” and “death.” Death is something that is an unknown existence for humans, and humans live their lives avoiding this thing called death. From this fact, unknown information may be considered to become the emotions that give rise to the unpleasant feeling that one is not safe. We think that, based on the instinctive desire of humans to change from an unpleasant state to a pleasant state, humans learn unknown information and thus make that information known to themselves.

In other words, we think that what humans do not learn reduces their safety. Thus they learn in order to be able to predict when an unpleasant state may be encountered, and they change to a pleasant state that increases their safety. Thus learning may be considered to change the feelings to pleasant which will lead to increasing ones safety. In this way, the time that one is thinking about ones own safety may be considered to be a factor that separates the pleasant and unpleasant emotions. Thus, the response is derived instantaneously, as mentioned earlier. In other words, by defining the state in which information is flowing smoothly as pleasant and the state in which it is not flowing as unpleasant, we think that it is possible to enable a robot to evolve in a more human-like and autonomous manner.

7 Conclusion

By means of these definitions, we have applied to a robot the fundamental human development principle of evolving based on emotions. And, by defining the many outstanding unresolved mysteries of the principles of human development in a visible way that we call the "smoothness of information," we have established the capability of emotion-based decision-making in the robot. As a result, we believe that we have achieved on the robot a system that takes into account the unification of reason with emotions and feelings. And we believe that we may have taken the first step towards achieving a robot capable of evolving autonomously of theconscious robot.

References

Chella, A., Manzotti, R. (2009). *Machine consciousness: a manifesto for robotics*, International Journal of Machine Consciousness, 33–51.

Donald, M. (1991). *Origins of the Modern Mind*, Harvard University Press, Cambridge.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, R. (1996). *Action recognition in the premotor cortex*, Brain 119, 3–368.

Holland, O., Goodman, R. (2003). *Robots With Internal Models: A Route to Machine Consciousness?*, Machine consciousness, 2–7.

Komatsu, T., Takeno, J. (2011). *A conscious robot that expects emotion*, International Conference on Industrial Technology (ICIT), IEEE, Auburn, USA

Kushiro, K., Takeno, J. (2011). *Development of a robot that cognizes and learns unknown events*, BICA, volume 233 of Frontiers in Artificial Intelligence and Applications, page 215–221. IOS Press.

Oudeyer, P. (2003). *The production and recognition of emotions in speech: features and algorithms*, International Journal of Human-Computer Studies, 157–183.

Takeno, J. (2011) *MoNAD structure and self-awareness*. Biologically Inspired Cognitive Architectures (BICA), Arlington, USA

Takeno, J., Inaba, K., Suzuki, T. (2005). *Experiments and examination of mirror image cognition using a small robot*, The 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp.493–498 CIRA, IEEE Catalog, 05EX1153C, ISBN0-7803-9356-2, June 27–30, Espoo, Finland.

Williams, R. (1987). *A class of gradient-estimating algorithms for reinforcement learning in neural networks*, Proceedings of the IEEE First International Conference on Neural Networks.