Consistency between Cognition and Behavior Creates Consciousness

Keita INABA, Junichi TAKENO
Science and technology, Meiji University
Ikuta, Kawasaki, Kanagawa, Japan

ABSTRACT

What is consciousness? Is it possible to create consciousness mechanically? Various studies have been performed in the fields of psychology and cerebral science to answer these questions. As of yet, however, no researchers have proposed a model capable of explaining the mind-body problem described by Descartes or replicating a consciousness as advanced as that of human beings. Ancient people believed that the consciousness resided in a Homunculus, a human in miniature who lived in the brain. It is no mystery that the ancients came up with such an idea; for consciousness has always been veiled in mystery, beyond the reach of our explorative powers. We can assert, however, that consciousness does not "live" in us, but "exists" in us. Insofar as the processes occurring inside the human brain are a product of the physical activity of the neurons that reside there, we believe that it should be possible to define consciousness systematically.

Keywords: consciousness, mirror neuron, qualia, robot, self consciousness, cognition, behavior

1. Introduction

Two developments may prove useful in unlocking the mystery of consciousness. The first is the discovery of mirror neurons. The mirror neuron [1] was discovered in the monkey brain by Professor Rizzolatti at the University of Parma, Italy. Clustered together within the human brain, mirror neurons fire not only when a person performs a certain action, but also when the person observes another performing the same action. Thus, a class of neurons functionally involved in both performing and acknowledging an action has been discovered. Studies using functional magnetic resonance imaging (fMRI) also identified regions with clusters of neurons equivalent to these mirror neurons (mirror systems). The discovery of the dual nature of mirror neurons attracted a great deal of attention. The neurons had two types of reaction characteristics, i.e., reaction in acknowledging so-called visual information and reaction related to kinetics, complex functions that remained unexplained by earlier paradigms. When scientists referred to the "premotor cortex" in the past, they assumed that each function of the brain was handled under a paradigm of "functional localization." Though the detailed function of mirror neurons is still unclear, the neurons are thought to assist kinetic learning (imitation) in the processes of imitating other people's actions and reading other people's psychological states (mind-reading). This close association between one's own acts and those of other people at the neuronal level suggests that information processing is carried out with a profound consciousness of one's self and others. Though not focusing on consciousness itself, Professor Yoshihiko Nakamura at Tokyo University designed a series of studies to replicate the function of mirror neurons by assigning a humanoid robot the task of imitating human beings through observation. [2] The second development of potential use in unlocking the mystery of human consciousness is the "mimesis theory" put forward by Merlin Donald. According to mimesis theory, human intelligence evolved during the first stage of a more than 3 million year evolution of the human mind and culture, a stage called "mimesis," while language is an extremely new phenomenon with a history of only about 500,000 to 300,000 years. [3] Donald's theory gains support from evidence corroborating that while spoken language probably dates no further back than 300,000 to 500,000 years, our human ancestors had already started hunting as groups and forming social groups with division of labor 1.5 million years ago. These social groups are considered to have developed without spoken communication. Communication before the birth of language is thought to have been achieved by imitation, or more precisely, "mimesis communication," an information process involving the creation and acknowledgment of signals freely. To utilize mimesis communication, what is generally considered an external model of communication, the information needs to circulate bi-directionally through the bodies or brains of the communicating participants. Thus, mimesis communication, i.e., imitation, plays a role in consciousness similar to the role of mirror neurons, from the viewpoint of consciousness of self and other. Another approach to consciousness that concerns us is the hierarchy chart of levels of consciousness proposed by Professor Naoyuki Osaka [4] of Kyoto University. (Fig. 1)

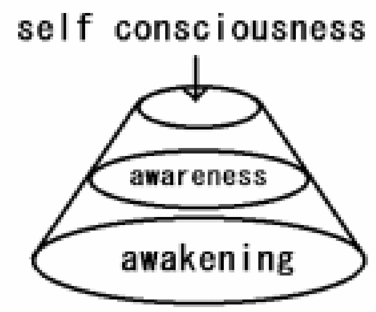

Fig.1 chart of levels of consciousness

In Osaka's hierarchy, consciousness is classified into three types, each of which is subdivided into three sub-hierarchies called awakening, awareness and self consciousness. In this chart, the upper hierarchy is assumed to have a deeper cognitive function. In this report, we explain these three types of consciousness using a new consciousness system we have propose based on our agreement with this concept. This report also devotes attention to "qualia," one of the so-called "hard problems" in consciousness study. Qualia are unique textures of sensations, such as those experienced upon perceiving the color red, the coldness of water, or the sweetness of sugar. Thus, qualia are in some way subjective senses that cannot be expressed in the amounts, quantities, units or symbols conventionally used to describe material systems. Though the subjective "I feel~" structure of a qualia seems impossible to replicate, the structure is produced through neuronal activity that can be quantified as a physical phenomenon. This is why qualia are regarded as a hard problem in the field of brain study. In this report we will attempt to provide a clear definition of consciousness based on the matters described above, and then illustrate the appropriateness of this definition. At the same time, we will propose a new model of consciousness in line with "cognition and behavior," using a consciousness system built up around our definition of consciousness.

2. Outline of consciousness system

In this report we define consciousness as "consistency of cognition and behavior." In order to replicate consciousness, we look at the "consciousness system." [5] We use this system to explain each function of the aforementioned consciousness levels proposed by Professor Osaka, and we use a robot to actually demonstrate this consciousness. To begin with, we will explain why we define consciousness as "consistency of cognition and behavior" by examining the path along which the consciousness develops and then elaborating our definition. Then we will attempt to prove that the three hierarchies mentioned above are indeed explainable. We will also explain those points that can be demonstrated by defining consciousness as "consistency of cognition and behavior".

2.1 Why consciousness was born?

With the discovery of mirror neurons and the development of mimesis theory, there can be little doubt that imitation learning relates closely to human evolution and the advanced development of intelligence. The authors of this report are of the opinion that imitation learning is connected to the germination of consciousness in human evolution. Human beings can tell the difference between self and others, as well as accurately recognize their own reflections and the reflections of others in mirrors. These abilities are crucial to the process of learning by imitation. It seems, firstly, that our ability to recognize these things allows us to collate the information necessary for imitation learning, and secondly, that this collation of information contributes greatly to imitation learning. That is to say, the authors hold that self consciousness of human beings was born from imitation learning. As will be described in greater detail later, imitation learning is assumed to be achieved through "consistency of cognition and behavior." Based on this idea, the authors define consciousness as the same process, "consistency of cognition and behavior." If we can develop a robot that has a "consciousness system" based on this definition of consciousness, we believe that it will be possible furnish the robot with an actual consciousness.

2.2 Composition of consciousness and explanation of each consciousness level

The diagram in Fig. 2 outlines the language function according to the consciousness system.
The composition of the consciousness system is explained here using the language function as an example, but we believe that this system can be applied to other behaviors in the same way. This consciousness system consists of a cognition system and behavior system. The cognition system mainly controls cognition. In this figure, sounds heard by the ears are learnt and recognized under the cognition system. The behavior system mainly controls behaviors in the same manner, and learning and action are controlled by moving the mouth and speaking words. We assume that the cognition system and behavior system are formed by almost the same neurons. This is an important point. In this consciousness system, however, cognition and behavior are learnt at the same time, or in series, as well as in cycles.

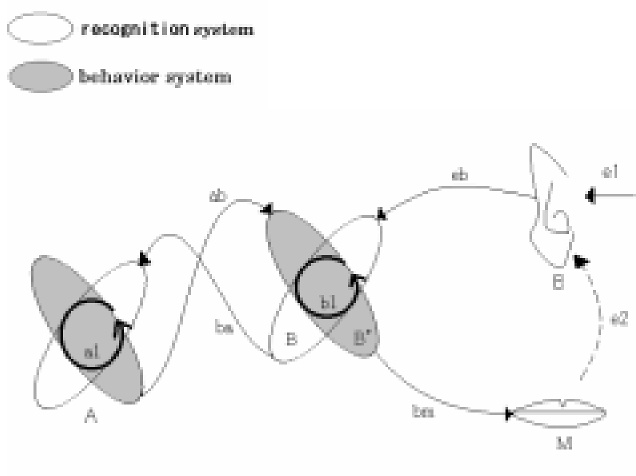

Fig.2 language function according to the consciousness system

2.2.1 Explanation of awareness in the consciousness system

The consciousness system can be explained using the phenomenon of awareness.
Awareness is defined as being when your attention is directed toward the external world, selectivity for stimulation or reaction based on the attention is recognized. We explain this status following the flow of information of language function in the consciousness system. In Fig. 2, when the ear E hears a certain item of language by e1, the language is understood by Cognition System B via eb. This information reaches a main part of consciousness System A incorporated in the ego via ba. This is the flow of cognition occurring when we hear a sound. When cognition is made in the cognition system we propose, the behavior of speaking the sound is simultaneously being learned. The circulation of information is performed according to a1 and b1. When a person takes the behavior of pronouncing a certain word, the command to pronounce is issued by a main part of consciousness System A, and then the command reaches Behavior System B’ via ab and is transferred onward through bm. The word is then pronounced by the Mouth M, the sound e2 is heard by the ear E, and the information is circulated. Furthermore, circulation of cognition is also carried out by a1 and b2. Awareness is the status of directing attention to the external world. This is actually the status described above, i.e., the cyclical status of cognition and behavior. In the status where learning of cognition and behavior is cyclical, we can take behavior and acknowledge the external world simultaneously through this circulation of information. This model also explains imitation. Since learning of cognition and behavior is performed both in series and cyclically, we can acknowledge the behaviors of others and learn them as our own behaviors right away.

2.2.2 Explaining self consciousness by the consciousness system

Self consciousness is defined as recursive consciousness, which is to say, the state of realizing what we do. In the explanation of awareness, we stated that this model has the function of imitation. As suggested by the mirror neurons or the mimesis theory described previously, this imitation function is thought to have a profound relationship with self consciousness. Imitation is the behavior of acknowledging the behavior of another person and replacing it as a behavior of one's own. Thus, we see that we incorporate the models of other people into our own models. This incorporation of external models makes up an inextricable part of self consciousness. Fig. 3 illustrates the function of moving the human hand using the consciousness system. A mirror is used to explain the concept of self consciousness more accessibly. Self consciousness can be explained in detail by following the flow of information.

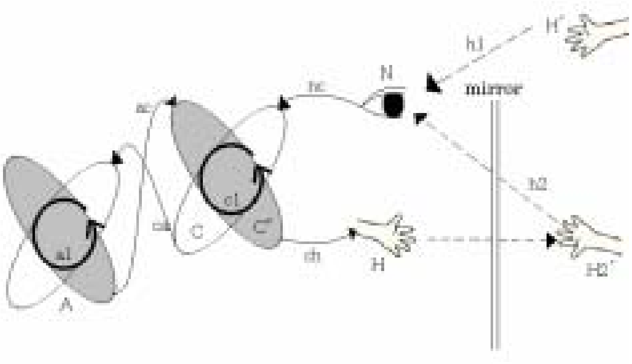

Fig.3 human hand function according to the consciousness system

The circulation in the learning of cognition is almost the same as that in the language function. When you try to move your hand, the behavior command is issued from A to Behavior System C' via ac, then goes through ch, whereupon the hand H is moved. This information circulates through a1 and c1 and influences the learning of acknowledgment. The flow of cognition is circulated normally by acknowledging the movement of another person's hand. The behavior of another person's hand h1 is acknowledged by the eye N and reaches Cognition System C via nc. Next, it reaches a main part of consciousness System A via ca. The information is circulated by a1 and c1 and connected to the learning of the behavior. If information is circulated by a1 and c1, the circulation of this information itself is nothing more than the incorporation of the model of another person, as described previously. The circulation connects the cognition of behavior by another person with one's own behavior, hence the process of reflecting another person's behavior in one's own behavior is thought to be identical to the process of incorporating the model of another person into one's self. Here, we assume that the behavior of another person's hand H' is the behavior of one's own hand H2' viewed in a mirror. Under this assumption, the model is indeed incorporated through circulation of a1 and c1, but it remains unclear whether it is incorporated as a model of another person's hand or of one's own hand. When information is circulated according to a1 and c1, the authors conjecture that we recognize the model as the model of another person when there is a large divergence between the circulating information, i.e., the acknowledged information and the information used to take the action, whereas we recognize it as a model of our own when there is only a small divergence between these two types of information. As an example, we can compare the cases of replacing behaviors taken by one's self or another person, as seen in a mirror. When replacing our own behavior seen in a mirror, it might be more appropriate to think that there is a very small difference in the information seen in the reflected behavior. If these conjectures hold, we might be able to explain self consciousness within the consciousness system the authors propose as the incorporation of our own models into ourselves, i.e., our understanding of the image shown in the mirror as being ourselves.

2.3 Qualia

The authors would also like to put forward several concepts regarding qualia. By our reckoning, a subjective act with selectivity or so called behavior itself is a qualia.
For instance, when we say, "I feel the color red, ” the color red itself is not a qualia; rather, we are acknowledging the subjective behavior based on the memory and learning of "I feel the color red" as a qualia. The problem of qualia is the problem of how subjectivity is created. If we say a qualia is "a behavior, ” or more precisely, "a memory of an behavior occurring due to an behavior, ” we believe that the problem of subjectivity can be solved since the behavior is subjective.

3. Experiments

This chapter describes experiments actually performed to create a consciousness system with a neural network (NN) using C language. We performed a simple simulation to investigate the performance of our NN, as a preliminary step towards the studying the evolution of language systems in an evaluation robot (Khepera) using the consciousness system. We actually taught Khepera to output the result calculated using an operator (AND, XOR, etc.) for an inputted 3bit value, then examined the result of learning at that time, and the number of learning times till learning reached the threshold. The BP(Back Propagation) method was applied for the NN used in this study. In order to simplify the problem, we considered the Cognition System and Behavior System as an NN formed respectively by 3 layers of 3 inputs, 3 outputs, and a mid layer. First of all we prepared an NN formed by 5 layers, with the Cognition System and Behavior System connected in series. Fig. 4 is a schematic diagram of the 5-layer NN. As the figure shows, the 3rd of the 5 layers is used both as output for the Cognition System and input for the Behavior System.

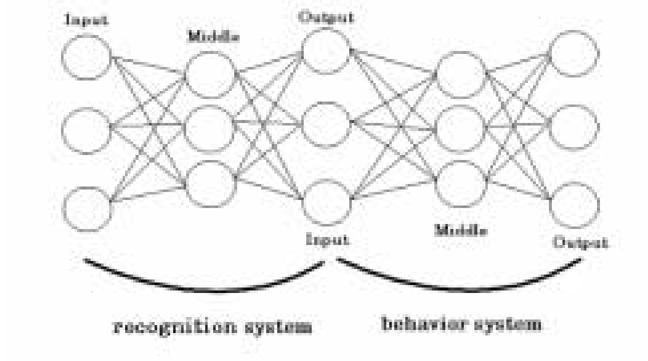

Fig.4 5 layers NN of consciousness system

Next, we added a circulation function to the 5-layer NN, converting it to a 6-layer NN with the Cognition System and Behavior System connected in series in the same manner described above.
Fig. 5 gives a detailed explanation. The 6-layer NN has the same contents as the 5-layer NN up to the 5th layer, but the

result of learning action obtained in the 5th layer (B) is given as input for cognition (A). Thus, the output of action is controlled in the 6th layer (C), and subsequently given as input for cognition again. As will be explained in greater detail in our ensuing description of control, threshold is set so that output of action is not circulated as input of cognition unless the learning reaches a certain level. This is because learning might fail to converge if an output value is used as an input value when learning is incomplete. In the case of actual human actions, an acknowledged action that differs greatly from the action we attempt to perform seems to be recognized as a failed or different action. If learning advances up to a certain point and we can perform the action better, we seem to acknowledge that we have performed the action and the learning efficiency of the action is promoted further.

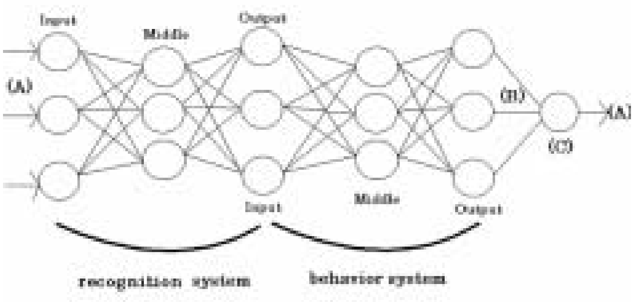

Fig.5 6 layers NN of consciousness system

Here we compare the learning result using only the Cognition System and Behavior System with the learning result of the 5-layer NN in which cognition and behavior are connected in series. The purpose of this experiment is to observe the variation in learning efficiency when we consider a consciousness system in which the Cognition System and Behavior System are united. In all the diagrams that follow, the x axis shows order (number of learning times) and the y axis shows errorfunc (square average error). Errofunc is the square average error of the teacher signal and output value, and learning continues until this value drops below a certain level (0.01).

Fig. 6.1 shows the learning result for XOR with a 3-layer NN.

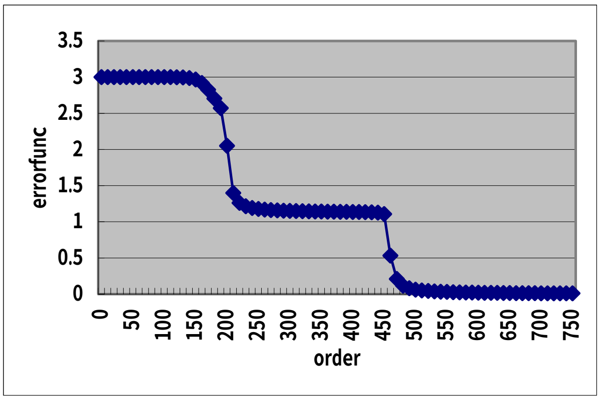

Fig6.1 result for XOR

As the figure shows, the learning was promoted and the order was 750 at completion of learning. Fig. 6.2 shows the learning result for NOT with the 3-layer NN.

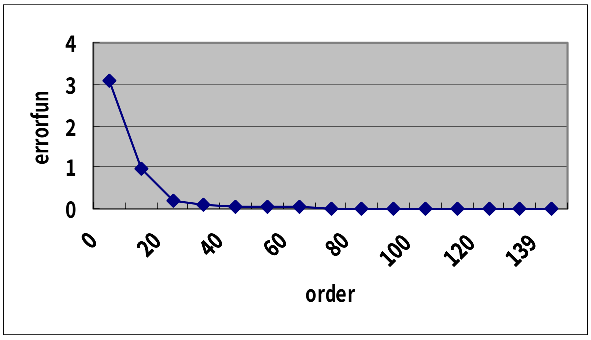

Fig6.2 result for NOT

Learning was promoted in the manner shown, and the order was 139 at completion of learning. We examined the learning efficiency when the learning was performed simultaneously under the consciousness system in which cognition and behavior were united, treating the learning of each XOR and NOT as cognition and behavior

Fig. 6.3 shows the result when XOR and NOT were learnt simultaneously under the consciousness system with the 5-layer NN.

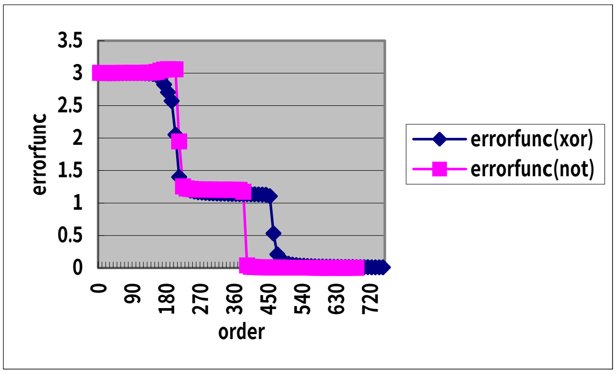

Fig6.3 result for XOR-NOT

XOR was learnt in the first 3 layers and NOT was learnt in the last 3 layers. The efficiency of learning NOT in the last 3 layers dropped as if it had been held back by the learning of XOR. Even if the learning of output in the first layers was imperfect, it was still sent to the last layers as output in the consciousness system. However, at order 750, the two types of learning dropped below the threshold in the end, and the number of learning times became the same as that when XOR was performed with the 3-layer NN. When we repeated the procedure with other basic operators (AND, OR, NOT, NAND), learning efficiency never dropped below that of the system in which cognition and behavior were separated, even under the consciousness system. This phenomenon is thought to have a high degree of generality, given that the theory of logical circuit design guarantees the composition of complex circuits up to a certain degree according to the combination of basic circuits. Based on the abovementioned observation, we can assume that learning efficiency will not drop if we use this consciousness system in a robot, whereas it will in the normal system in which cognition and behavior are performed separately. Next, we looked at the learning result when NOT was learned in the first 3 layers and XOR was learned in the last 3 layers, as in the case with the operation NOT-XOR, using the 5-layer NN. Since the XOR result learned in the first 3 layers held back the learning of NOT in the previous experiment, we performed this next experiment to see how the learning efficiency related to the

agreement between the value used for input of the last 3 layers and the learning result of the first 3 layers.
Refer to Fig. 6.4.

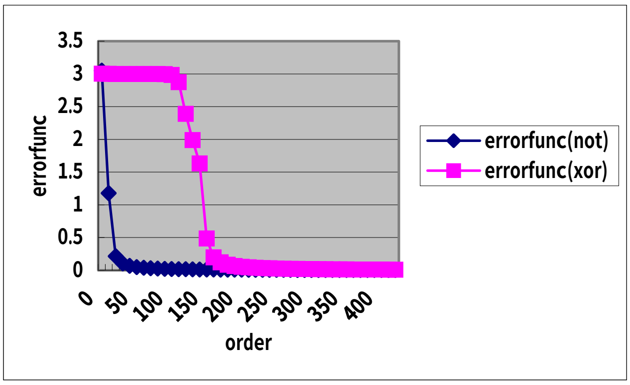

Fig6.4 result for NOT-XOR

In this case as well, the learning of NOT in the first layers held back the learning of XOR in the last layers. However, the XOR learning showed a higher learning efficiency, since the speed of NOT learning was faster than that of XOR learning. When these two learning items reached the threshold and finally finished, the order was 427. Learning efficiency was higher than that in the former system in which cognition and behavior were performed separately.
We can assume that learning efficiency was increased, since the learning result for NOT used for the initial input of XOR happened to be a value favorable for increased efficiency. When we repeated the same procedure with other operators (NOT-NAND), the learning efficiency increased several tens of percent. Given this increase in learning efficiency and the absence of any declines in learning efficiency when the abovementioned 5-layer NN was used, we can assume that the learning efficiency will not decrease in this case, either. According to these two experiments, we can assume that the learning efficiency will not drop with the use of the author's consciousness system in which cognition and behavior are united, whereas it will with the use of the former system in which cognition and behavior are separated. Furthermore, we have also learnt that we can increase learning efficiency according to the learning method. It was clarified that simultaneous learning in which cognition and behavior are connected in series presents no problems in terms of learning efficiency.
Next, we experimented with a 6-layer circulating consciousness system identical to the 5-layer system up to the first five layers. Here, the output result of "learning an action" was not circulated from the 6th layer to the 1st layer for use as input for "learning of cognition". Instead, we set a 7th layer and examined learning efficiency, treating the learning from the 1st to the 7th layer as 3 NNs.
The system is the same as that shown in Fig. 4 up to the first 5 layers. Between the 5th and 7th layers, the value is converted in order to use the output result from the learning of an action as an input for the learning of cognition. As shown in Fig. 5, the value is actually controlled at the 6th layer and circulated back to the 1st layer. But in order to compare the learning efficiency according to threshold of the control, we did obtain the number of learning times by circulating the 7th layer as output. In the formula used for this experiment, the even number bit was "NOT" in the first NN, the odd number bits were "NOT" in the second NN, and all of the bits were "NOT" in the third NN. The value is assumed to return to the original input by three calculations when circulating. Thus, the output of action and input of cognition become the same. Fig. 6.5 shows the learning result.

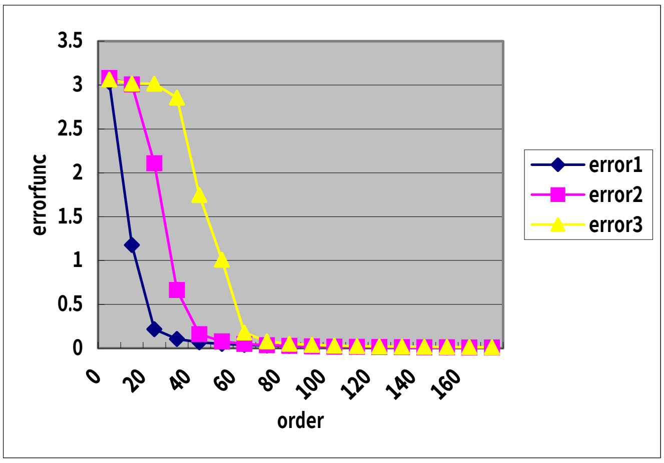

Fig6.5 result for 7 layers NN

Learning was promoted in the manner shown in the figure, and it converged at order 171. Errors 1, 2, and 3 in the graph show the degree of learning progress of the NN from the 1st to 3rd layers, from the 3rd to 5th layers, and from the 5th to 7th layers, respectively.
Next is circulation from 6th layer to the 1st layer. Since learning is considered not to converge when it is deficient in circulation, we set a threshold to start circulation after learning has been promoted above a certain level. Let's look at the learning efficiency when the circulation threshold is 1.2. (The circulation threshold is the difference between output at the 6th layer and the teacher input at the 1st layer.)
Fig. 6.6 shows the learning result when the circulation threshold is 1.2.

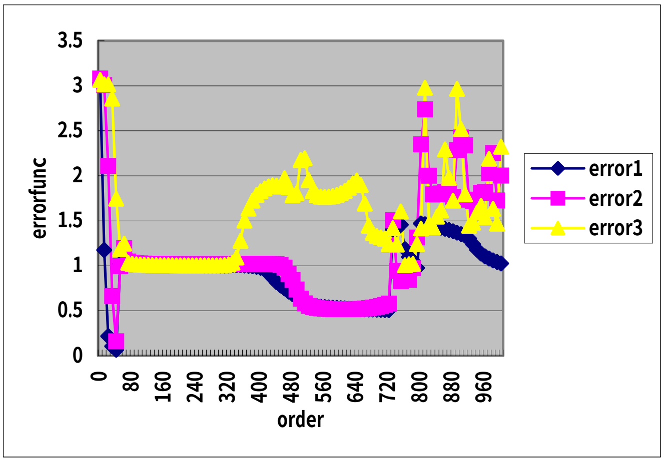

Fig6.6 result for 7layers NN(threshold1.2)

Here the convergence stopped in the middle, even though the learning was promoted from the start. We circulated up to order 1000, but still obtained no convergence of learning.
Based on the above result, we dropped the circulation threshold to 0.8. Here the convergence started earlier than it did when the circulation threshold was 1.2, but the final order became 350. The learning efficiency here dropped below the learning efficiency in the case of no circulation. Next, we dropped the circulation threshold to 0.4. Here, the convergence was promoted steadily and learning converged at order 163. Thus, the learning efficiency had risen above that in the case of no

circulation.
Through this experiment, we elucidated the difference in learning efficiency based on the circulation threshold of the consciousness system. When the circulation was initiated before the start of learning, the learning convergence dropped dramatically. However, we also learned that the learning efficiency could be increased by controlling the circulation. From the viewpoint of learning efficiency, our experiments revealed no decreases in function with the actual use of our consciousness system in the Khepera robot.

4. Prospects for the future

Henceforth we plan to use the consciousness system created in this report in the advanced robot Khepera and hold experiments with a robot that can perform self cognition. The Turing test was formerly used as a method for judging whether artificial intelligence had been realized. Though we have no plans to clarify whether the Turing test was right or wrong here, we should mention that no other criteria equivalent to this Turing test has yet been defined in studies of consciousness.
Still lacking a means to determine whether a robot has consciousness, we plan to conduct an experiment equivalent to the Turing test for consciousness study. Using several robots and a mirror, we will construct a dialog in which robot A, the main figure of the experiment, converses with other robots and robot A', its own a reflection in the mirror. Our object will be to determine whether robot A can distinguish between the conversation with robot A', its own reflection, and conversations with the other robots in the experiment; in other words, whether robot A can recognize itself.
Concretely speaking, we plan to construct a language system in the advanced robot Khepera using the consciousness system, and perform imitation learning of the language function to investigate whether a robot can discover the abovementioned difference. The language system will be composed using IOtalet to emit light from the Khepera itself, treating the bit row of lights as the pronouncing activity to be recognized by the other Khepera.

5. Conclusion

Consciousness study is rapidly being developed, with a good many studies underway and diverse opinions being presented. Within this scenario, it is very important to consider consciousness from the point of view of brain system theory.
To define and clarify consciousness in this report, we examined concepts on the germination of consciousness implied by the discovery mirror neurons and their dramatic role in information processing of the brain, as well as the development of mimesis theory and its application to human evolution.And by using the definition of this consciousness showed that explanation of human's various consciousness phenomena could be performed.
Next, we examined a consciousness system that replicated consciousness, and attempted to explain the three kinds of human consciousness using the consciousness system.
Finally, we created actual consciousness system and performed simple simulation experiments. As a result of these experiments, we found that this consciousness system can be used without problems, and achieves learning efficiency superior to that obtained with the former robot system. We also established that learning efficiency could be improved over the former system when the appropriate learning methods were applied. Experiments were performed for each basic operator in order to examine the versatility of the consciousness system.
Finally, as a prospect for the future, we proposed a new Turing test for consciousness.

Reference

[1] Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) “Action cognition in the premotor cortex.” Brain 119: 593-60

[2] Toshihiko Nakamura. “Mirror Neurons Models which absorb Difference of Body Conditions between Teacher and Learner.” Robotics Society of Japan 20th (2002)

[3] Merlin.Donald.: “Origins of the Modern Mind” Harvard University Press, Cambridge, 1991

[4] Naoyuki,Osaka: “Brain and consciousness.” Asakura Press (1997)

[5] Junichi Takeno.: Protected Intellectual Property, “ How to make a consciousness of robot”, Application No.189760, Japan Intellectual Property Association, 11th June 2002.

[6] Junichi Takeno: New Paradigm “Consistency of cognition and behavior”(The Best Paper of Methodology), International Conference on Communication and Control Technologies CCCT2003, The International Institute of Informatics and Systemics(IIIS), Proceeding Vol.I, ISBN 980-6560-05-1, pp.389-394, 2003-8.

[7] Junichi Takeno, K. Inaba, T. Suzuki: “Research related to imitation behavior using a consciousness machine”, International Conference on Communication and Control Technologies CCCT2004, The International Institute of Informatics and Systemics(IIIS), Proceeding Vol.I, ISBN 980-6560-17-5, pp.268-273, 2004-8