Apparent "free will" caused by representation of module control*

Natsuki Oka
Matsushita Research Institute Tokyo, Inc., Tokyo, Japan

1. Introduction

Constructing conscious level processing with neural networks has been one of the most exciting challenges in cognitive science. However, there remains a wide gap between conscious level processing and low level processing which is realized with neural networks. When people think over the functions of consciousness, they tend to take notice of higher level functions, such as language comprehension and self-awareness; however, I believe that in order to fill the gap, we should begin by studying consciousness in its earlier stages of evolution or development, in which stages no symbolic representation is processed.

I therefore tried to consider the examples such as:

- Behavior of infants who can not or hardly speak yet. Some of the examples are:
  1. My daughter took the remote controller of a television, directed it against the television, and had fun with pushing some buttons on it. [11 months]
  2. She handed a package of sweets to me, because she could not open it. [1 year and 2 months]
  3. She took me out of a room, then she alone came back to the room and climbed a chair. It seems that she did not want to be stopped from or be assisted in climbing. [1 year and 4 months]
- Behavior of animals as in Dawkins (1993).

In the first example of the behavior of my daughter, it seems that she did a novel action by imitating that of her parents; and in the second and third examples, it seems that plans of actions were made and she could cope with new situations. Imitation and planning are two basic functions that generate novel action patterns without a long training period, which I believe is the first distinctive feature of conscious level processing.

The second distinctive feature is to be able to generate novel thought patterns immediately by imitation or planning of thought patterns. It is important to note that the newly acquired thought patterns can be used for generating further new action or thought patterns, and that the circular structure of this kind will give rise to the emergent properties of our model.

To sum up, the characteristics of conscious level processing are:

1. quick composition of novel action patterns, and
2. quick composition of novel thought patterns.

In the next section, I propose a connectionist model of conscious level processing which can immediately produce new action patterns and new thought patterns. Then I try to relate the phenomenon of apparent "free will" with the characteristics of our model, and the last section contains conclusions and future work.

2. A connectionist model of conscious level processing

Let us suppose that there exist neural network modules each of which has elemental functions such as recognition and prediction. For example, some of the modules may infer state changes such as

My father has a package of sweets. → The package is open.

and this kind of association would be learned from seeing similar state changes several times.

Complicated processing can be realized by combining these modules. In order to change the combination freely and variously, a place or places where the inputs/outputs of the modules gather seems necessary, which I call working memory (see Figure 1). Each input/output has a gate, and the states of the gates are controlled by the modules, thus the information flow among the modules, that is, the combination of the modules is controlled by the modules by means of gates.

Note that knowledge which includes variables can be represented as connected modules. For example, a formula:

$$p(X) \leftarrow q(X, Y) \,\&\, r(Y),$$

where p, q and r are predicates, and X and Y are variables, is represented as module combination shown in Figure 2.

The model generates novel action patterns by imitation and planning (Characteristic 1). I assume that imitation and planning are built-in thought patterns, and that they are realized by module combination. Imitation is a thought pattern consists of recognition and action. As an example, a schema of visual imitation is shown in Figure 3b. The specific combination of the recognition and the action module is formed by a gate control module which is fed with 'imitation' input (Figure 3a).

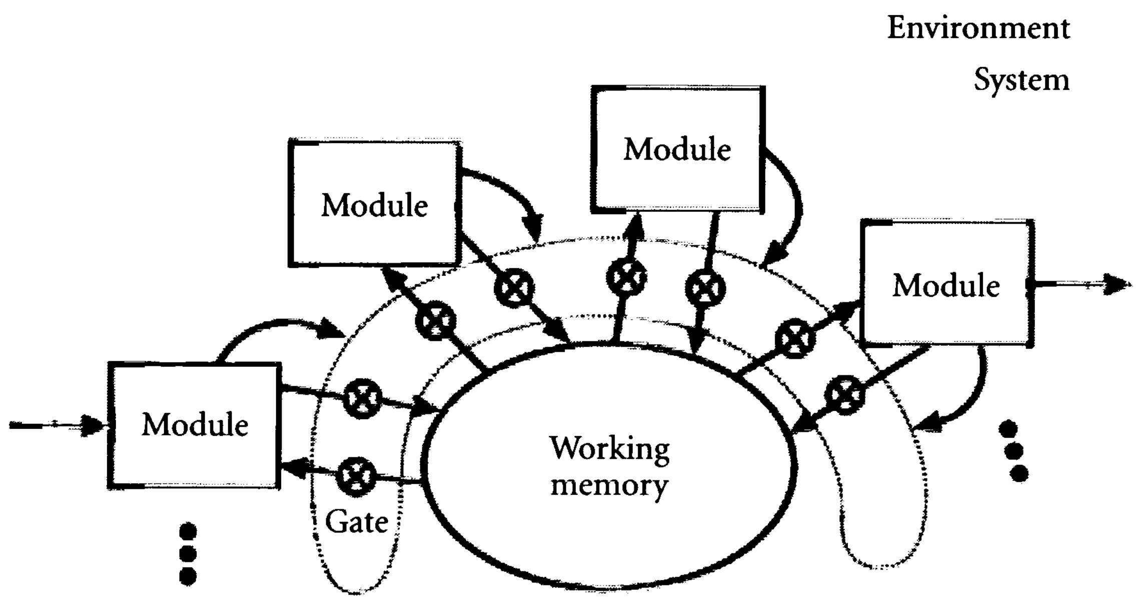

Figure 1. A connectionist model of conscious level processing. Complicated processing can be realized by dynamically changing the combination of modules.

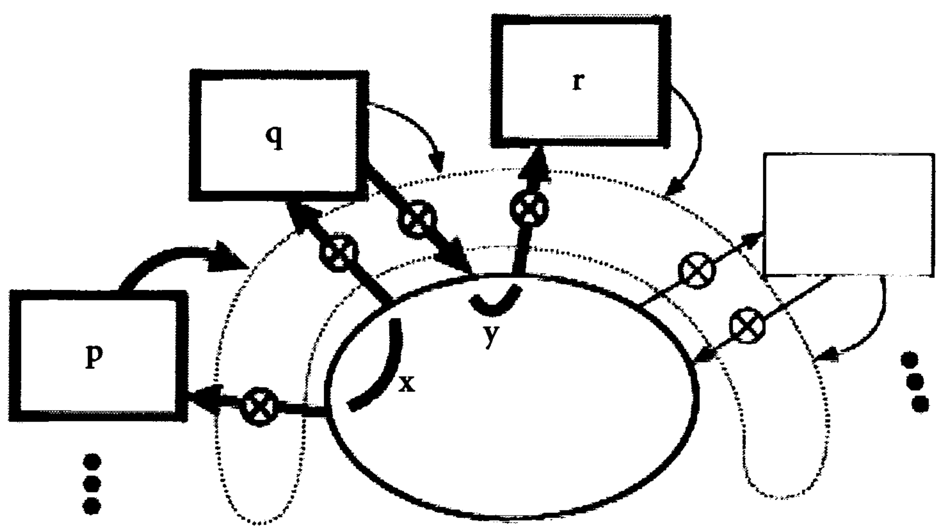

Figure 2. Knowledge that includes variables can be represented as connected modules. This module combination corresponds to a formula: $p(X) \leftarrow q(X, Y)$ & $r(Y)$.

Planning is realized by combining an action candidate generation module, a prediction module, an evaluation module, and so on (Figure 4b). The specific combination of the modules is formed by a gate control module which is fed with 'planning' input (Figure 4a).

Moreover, new combinations of modules, that is, new thought patterns, can

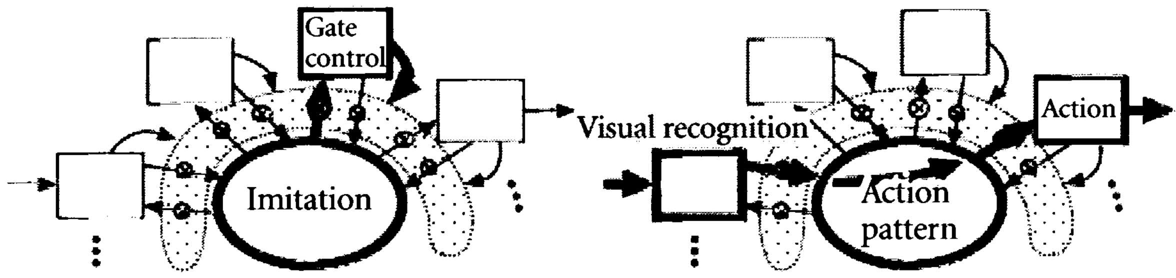

Figure 3. A schema of imitation. (a) A gate control module forms a module combination for imitation. (b) Imitation enables immediate acquisition of new action patterns.

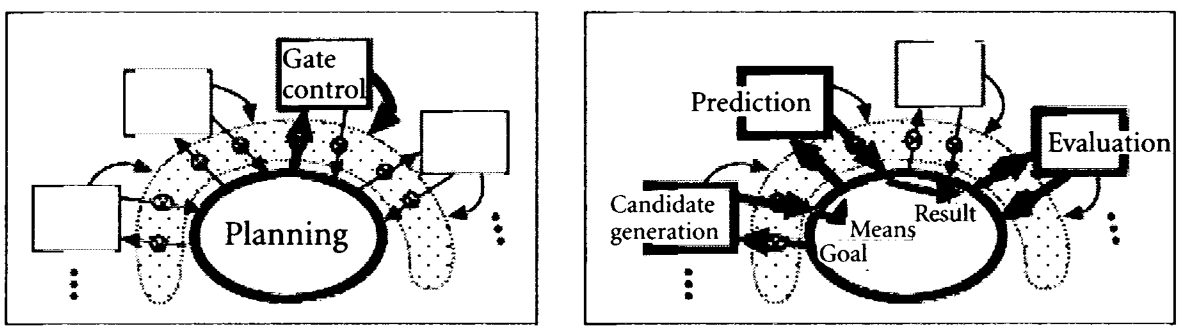

Figure 4. A schema of planning. (a) A gate control module forms a module combination for planning. (b) Planning enables quick acquisition of new action patterns.

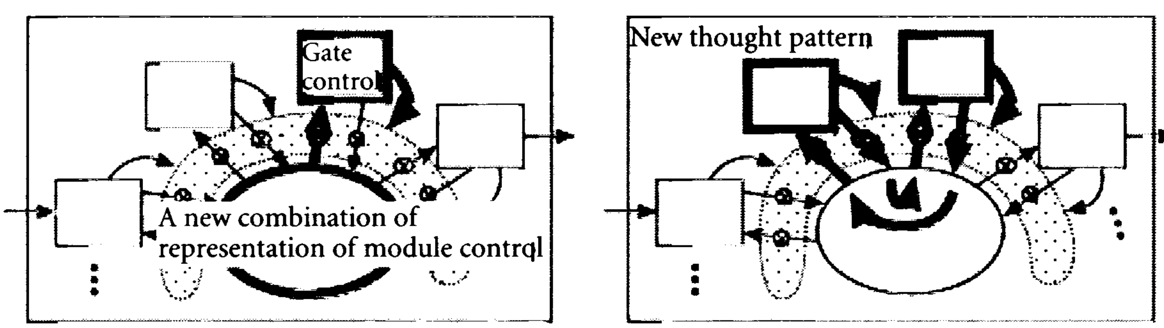

Figure 5. A schema of metalearning. (a) A new combination of control representations, which is quickly made by imitation or planning, in working memory composes a novel thought pattern. (b) The novel thought pattern can be used for generating further new action or thought patterns.

immediately be produced (Characteristic 2) by providing working memory with new combinations of representations of module control. See Figure 5 for illustration. The combinations of control representations can be produced by imitation or planning.

Not only built-in thought patterns such as imitation and planning, but also the newly acquired thought patterns can be used for generating further new action or thought patterns, which makes conscious level processing very powerful and autopoietic (Figure 6).

3. Apparent nondeterministic free will and the model

As is described in the previous section, combinations of the modules which correspond to thought patterns are represented in the working memory and they can be composed by imitation or planing, which means that the explicit control information exists, and that new representations which correspond to new combinations of modules can be produced by itself. These characteristics may relate to apparent nondeterministic "free will".

Another possible cause of apparent "free will" might be the existence of the modules which have the following functions:

Module A If the goal is to feel "free will", output some executable action candidates (for example, raising right hand and raising left hand) to working memory, and activate the input gate of **Module B.**

Module B Given some action candidates as inputs, select any one of them (this selection is done automatically and is not aware of), and activate the gates of

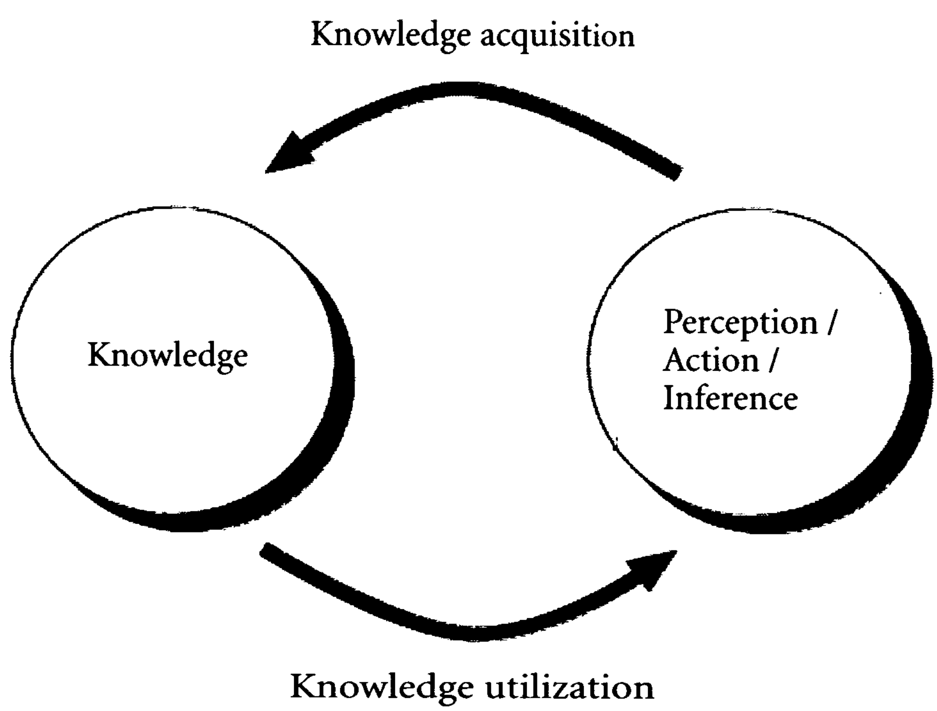

Figure 6. Autopoiesis. Perception, action or inference based on acquired knowledge makes further knowledge acquisition more efficient.

modules for the selected action.

Module C If the goal is to generate an explanation, and if the cause of the selection cannot be found, output the most plausible explanation, "the selection is done by my free will", to working memory.

Note that typical processing on the conscious level such as rehearsal or self-monitoring also can be described as a combination of the functions of some modules like the above example, and that the central control is not necessary for it.

4. Conclusions and future work

This paper proposed an idea of connectionist model of conscious level processing. I believe that this model will provide a useful insight into studying emergence and development of intelligence and consciousness.

Based on the proposed idea, we are going to build a knowledge acquisition system modeling after development of infants. We believe that the setting shown in Figure 7 is essential to development of cognition. Both a knowledge acquisition system and a person can act in their environment, and can recognize the environment. The person plays a role of a parent, and communicates with the system by language, gesture, and so on.

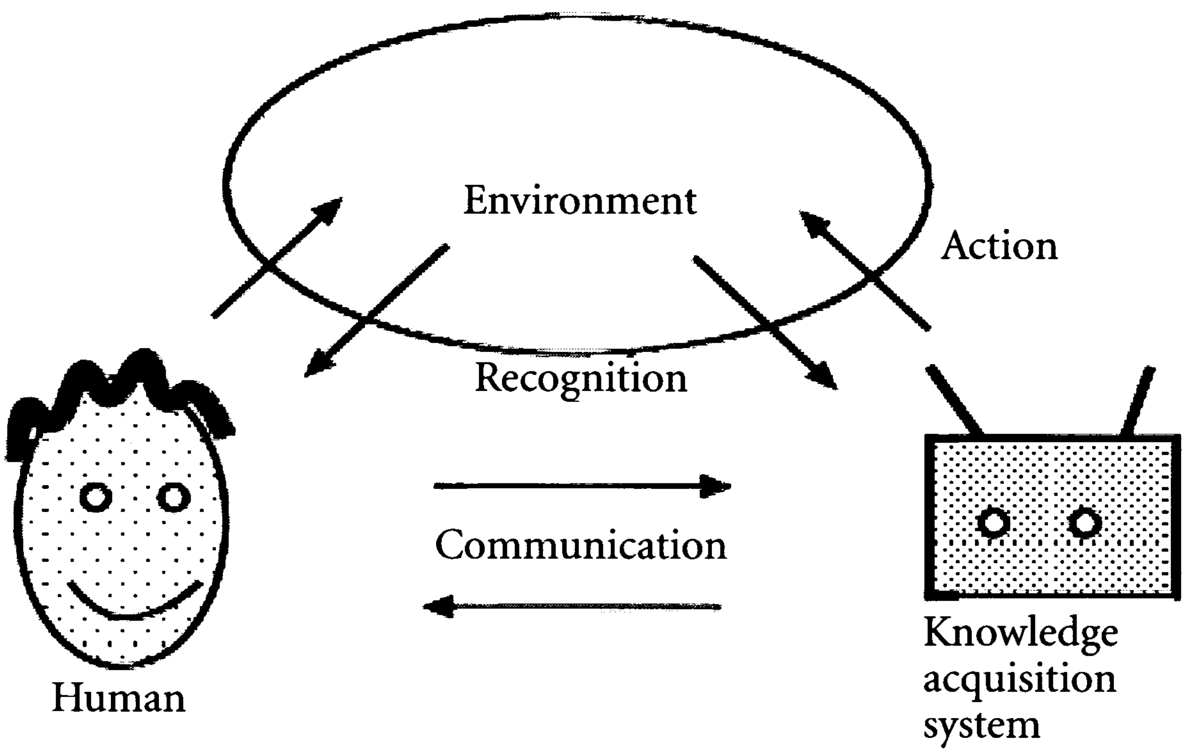

Figure 7. Knowledge acquisition setting. Interaction between a knowledge acquisition system, a human and their environment is essential to knowledge acquisition.

Although the setting of this kind is usually realized by building mobile robots which act in the real world, we are going to use a virtual world as the environment in order to minimize difficult problems of recognition and action in the real world, and to concentrate on the studies of knowledge acquisition. In our setting, the knowledge acquisition system does not have any physical body; however, it interacts with its environment and a person, and learns how to act and communicate in the environment, that is, it learns to recognize and predict input signals, forms concept hierarchies, and finds relations among signals in different sensory modalities.

Another important issue of this plan is to collect a large amount of interaction examples for learning. We are going to make it practicable by designing an enjoyable environment.

Note

* Thanks to Koji Morikawa, Takashi Omori, Mutsumi Imai, Kazuo Hiraki, Hidetosi Sirai, Masumi Ishikawa, Itsuki Noda, Hiroshi Nakagawa, Jun Tani, and many colleagues of Matsushita Research Institute Tokyo, Inc. and Matsushita Electric Industrial Co., Ltd. for valuable discussions and comments.

References

Dawkins, Marian S. (1993). *Through our eyes only?: The search for animal consciousness.* New York: W.H. Freeman.

Kawato, Mitsuo (1997). Bidirectional theory approach to consciousness. In M. Ito, Y. Miyashita and E.T. Rolls (Eds), *Cognition, computation and consciousness.* Oxford: Oxford University Press. (pp. 223–248)

Mathis, Donald W. and Mozer M.C. (1995). On the computational utility of consciousness. In G. Tesauro, D.S. Touretzky and T.K. Leen (Eds), *Advances in neural information processing systems 7.* Cambridge: MIT Press. (pp. 11–18)

Nakashima, Hideyuki, Noda, I. and Handa K. (1996). Organic programming language GAEA for multi-agents. Proceedings of ICMAS-96.

Oka, Natsuki (1991). Hybrid cognitive model of conscious level processing and unconscious level processing. IEEE International Joint Conference on Neural Networks.

Omori, Takashi. and Mochizuki A. (1995). PATON: A model of context dependent memory access by an attention mechanism. In R. Moreno-Diaz and J. Mira-Mira (Eds), *Brain processes, theories and models.* Cambridge: MIT Press. (pp. 134–143)

Tani, Jun and Nolfi S. (1998). Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. In R. Pfeifer, B. Blumberg, J.A. Meyer and S.W. Wilson (Eds), *Proceedings of the fifth International conference on simulation of adaptive behavior.* Cambridge: MIT Press. (pp. 270–279)