

Designing a Model of Consciousness Based on the Findings of Jungian Psychology

Toshiki Watanabe^{*1}, Hiroyuki Kameda²

¹Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology, Tokyo, 192-0982, Japan

²School of Computer Science, Tokyo University of Technology, Tokyo, 192-0982, Japan

ARTICLE INFO

Article history:

Received: 15 July, 2021

Accepted: 10 October, 2021

Online: 23 October, 2021

Keywords:

Model of consciousness

Jungian psychology

Artificial Intelligence

ABSTRACT

As artificial intelligence (AI) develops, it is expected that humans and AI will become more closely related than now. At the same time, however, the more closely humans and AI are related to each other, the more clearly they will face a moral dilemma, i.e., artificial intelligence will face a moral dilemma. To solve the moral dilemma problem, AI should understand and take into account human values and ethics. From this point of view, we designed a consciousness model based on Jungian post-psychological notes. As a result, we found that in order to implement the model of consciousness on a computer, it is necessary to design it with a structure similar to that of human beings, referring to human structures in various fields.

1. Introduction

This paper is an extension of work originally presented in SCIS-ISIS 2020 [1]

1.1. The social value of artificial intelligence technology

Artificial intelligence (AI) is currently being used in a variety of fields, including medicine, weather, and robotics. It has become an indispensable part of our lives, and its use is still being actively researched.

It is expected to evolve further in the future and will be applied and used in even wider fields than at present. According to an announcement by Japan's Ministry of Internal Affairs and Communications (<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd113230.html>), the use of AI is expected to lead to more efficient operations and the development of new products and business models. It is also expected to have the effect of creating innovation by freeing humans from cumbersome tasks and enabling them to focus on creative tasks that only humans can perform. Several research organizations have analyzed the effect of AI on productivity. For example, according to an analysis by Accenture, the labor productivity, which is defined as (results produced or amount of value-added) / (quantity of labor), e.g., number of workers or workers * working hours, in Japan is expected to increase by 34% in 2035 compared to the baseline. The value of AI will become even higher in the future.

1.2. Current Status and Trends of Artificial Intelligence

In recent years, AI has become so prevalent in our society that we can ask so-called voice assistants such as Alexa (Amazon), Siri (Apple), and Google Assistant to tell us today's weather, or help input a destination into car navigation or show us the route on a map. In addition, there is a lot of research that focuses on interaction with humans, such as mentioned in [2] and [3], where robots and androids communicate with humans to heal and support them. In the future, humans and AI will be more closely related to each other.

As humans and AI become more closely involved, some of the decisions that have traditionally been made by humans will be delegated to them. For example, in the area of automated driving, Waymo currently commercializes Level 4 automated driving in the U.S., and Japan has also released a roadmap for fully automated driving by 2030 (https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20200715/2020_roadmap.pdf), while Honda LEGEND (<https://www.honda.co.jp/LEGEND/>), which is currently Level 3 automated driving, is currently being sold to consumers.

However, with the development of these technologies, AI is facing the problem of moral dilemma. A thought experiment of Massachusetts Institute of Technology's Moral Machine (<https://www.moralmachine.net/>) exposes problems like the trolley problem to gather human perspectives on how self-driving cars should behave. We believe that this kind of problem is an important issue that must be considered in the future of AI.

*Corresponding Author: Toshiki Watanabe, g212003578@edu.teu.ac.jp

1.3. Position of this study

In this study, we hypothesized that if humans were to delegate decision-making to AI in the future, they would be able to make better and more flexible decisions by understanding and taking into account human values, ethics, and local culture and customs than now. We hypothesized that if the conscious and unconscious minds that control self-control could be realized on a computer, it would be able to understand and take into account human values, ethics, and local culture and customs. Self-control is considered to be important for this, and we thought that the relationship between the conscious and unconscious mind, which controls behavior and judgment, makes this possible. Therefore, we decided to model the conscious and unconscious as a decision-making system for AI and reproduce it on a computer for verification.

Therefore, to realize the conscious and unconscious on a computer, we designed a model of consciousness based on the findings of Jungian psychology.

2. Related works

2.1. The treatment of consciousness in the field of psychology

Freud is one of the first persons to discover the unconscious. He likened the human mind to the tip of an iceberg and divides conscious into three layers: the consciousness, the pre-conscious which can be made conscious, and the unconscious which cannot be made conscious, He also stating that repressed desires are pushed into the unconscious.

In his work, the author described the structure of the mind as consisting of the *es*, which is a mass of instincts, the *superego*, which is responsible for the protection of ethics, and the *ego*, which regulates them and guides conclusions to action [4].

The authors did not break down the human being structurally, as his study was centered on the theory of the purpose, which states that the individual mind cannot be divided and acts according to its present purpose [5].

The authors divided human conscious into the *consciousness*, the *personal unconscious*, and the *collective unconscious*.

Consciousness is that in the sense of being aware, and so on., while unconsciousness refers to other things.

"Many of the difficult perceptual computations, such as invariant face recognition or speaker-invariant speech recognition. that were recently addressed by AI correspond to nonconscious computations in the human brain. [...] Many of these C0 computations have now been captured by AI, particularly using feedforward convolutional neural networks (CNNs)."

The personal unconscious is the consciousness that has been suppressed by experiences and can be viewed as things that we are no longer conscious of, but which we can become aware of if we want to.

The collective unconscious is the unconscious that is common to all human beings and cannot be conscious.[6]

In Jungian psychology, the unconscious accounts for about 90% of the total, with the collective unconscious accounting for a large percentage of the total.

He was also the first person to use the term *complex*, as it is commonly referred to use in an even broader sense than its general scope.

In a recent study on the role of the conscious and unconscious mind the authors in [7],

"(1) Recent findings in social psychology research indicate that many human behaviors, including self-control, depend on automatic processes that do not require consciousness (2) We examine the relationship between conscious and unconscious processes in self-control and propose effective self-control in which the two processes work together. The purpose of this study is to examine the relationship between conscious and unconscious processes in self-control and to propose an effective self-control system in which the two processes work together. The role of conscious processes seems to be to form new coalitions through abstract mental editing. [...]One of the important functions of conscious processes was also to make unconscious self-control more effective. Thus, it is only by assuming the collaboration of the two that the essential functions of unconscious and conscious processes can be seen."

Through experiments, he came closer to elucidating the roles and functions of the conscious and unconscious.

2.2. Treatment of consciousness in the medical field

In the field of medicine, consciousness is often treated as a level of conscious or arousal. According to the website(<https://bsd.neuroinf.jp/wiki/%E6%84%8F%E8%AD%98>) ,it says

"The level of consciousness or arousal is the state of awareness. The level of consciousness is the highest when you are awake and clear-headed, lower when you are sleepy and fuzzy-headed, or when you are asleep and not dreaming, or when you are under deep anesthesia."

Typical scales for the level of consciousness include the Grady Coma Scale. In [8] the authors, the authors say

"The grade I patient is only slightly confused. The grade II patient requires a light pain stimulus (such as a sharp pin tapped lightly over the chest wall) for appropriate arousal, or may be combative or belligerent. The grade III patient is comatose but will ward off deeply painful stimuli such as sternal pressure or nipple twist with an appropriate response. The grade IV patient reacts inappropriately with either decorticate or decerebrate posturing to such deeply painful stimuli, and the grade V patient remains flaccid when similarly stimulated."

Consciousness in neuroscience and neurology deals with both psychological and medical consciousness. In [9] the authors discussed how to make computers conscious from the perspective of neuroscience. They argue that

He also claims that in other words, much of what AI does today, such as face and voice recognition, corresponds to unconscious computations in the human brain. They also claim

In other words, he claimed that if the Unconscious processing (C0) can be equipped with Global availability (C1) and Self-monitoring (C2), the machine will behave as if it has conscious.

2.3. Related studies in the field of engineering

In related research in the field of engineering, models of the mind and emotion have been actively studied.

"We contend that a machine endowed with C1 and C2 would behave as if it were conscious—for instance,"

In [10], the author builds a model that is aware of the whole picture of the mind and is hierarchical at five levels, from raw data to complex concepts, and can integrate low-level signal processing with high-level symbolic processing, which AI is good at. They built a model called the "Aesopworld" and implemented it in the Aesopworld project.

In [11], the author built a model of the mind based on the passive consciousness hypothesis, dividing it into two major parts: the conscious system and the unconscious system, assuming the passivity and illusion of consciousness. Thereby he proposed a method of conceiving the mind that has clear boundaries and is feasible, unlike conventional cognitive models.

In [2], the robot used cold reading and other techniques in communication to read the minds of customers and use them in sales. In [3] the authors adopt a model of personality to "heal" people.

These can be said to be equipped with unique models to communicate with humans.

In [12] the authors analyzed the trends in incorporating emotions into AI and what people need to be aware of when building models. People would like to create a foothold for clarifying the nature of emotions, which have been considered ambiguous, by incorporating knowledge about emotions from different fields such as psychology, neurology, and cognitive science, and other emotion research using a constructive approach. The reason why people can't reach the essence of emotions is probably that we can't abolish naive psychology and that the emotion research in various fields is conducted in different ways. He pointed out that it is important to abolish naive psychology to build a model of emotions and the mind.

2.4. Position of this study

In this research, we focus on being able to reproduce human consciousness on a computer. This is because if we can reproduce human consciousness on a computer, we will be able to systematically explain what people think and how they behave. And we believe that this will be useful not only in the field of AI but also in various other fields such as psychology and brain science.

3. Consciousness model in this study

In this study, we modeled consciousness in psychology, while including ethics, culture, and the human thinking and judgment, www.astesj.com

and aimed to construct a model of consciousness as a decision-making system, so that AI with the model can understand and take into account their values, ethics, and local culture and customs.

Psychology remains at a conceptual level and is difficult to implement directly in a computer, but it most clearly represents the role of consciousness and human behavior. On the other hand, research in the field of engineering has concrete models and roles to be implemented in computers, but they do not represent human behavior. In addition, the role of consciousness and human behavior was clear in the field of brain science, although the level of conscious in the medical field is different in meaning from the consciousness dealt with in this study. However, we still could not find a model that could be immediately implemented in a computer.

Compared to previous studies, my research is based on psychology, which allows me to design a model that eliminates naive psychology as much as possible, and by dividing the functions of consciousness, I can create a more detailed and clear model for designing programs.

The psychology we refer to when modeling is Jungian psychology, which divides the flow of thought into the conscious, personal unconscious, and the collective unconscious, and explains the role of them in particular detail.

In [9] the authors found that it is effective to divide the functions of consciousness, so our model in this study is also designed by dividing the functions to some extent to clarify the roles of each process.

3.1. Definition of our Consciousness Model

We define the consciousness model as a decision-making function that selects action based on ethics, culture, and values accumulated through personal experience.

3.2. Requirements for an Consciousness model

The model of consciousness will be set to meet the following requirements.

- To fulfill the roles of consciousness, personal unconscious, and collective unconscious in Jungian psychology.
- Input and output are based on the human body.
- The model has memory retention and storage functions.
- The model can be reproduced as a program on a computer.

3.3. Pre-design of Consciousness Model

Consciousness in Jungian psychology is regarded as a system, and the functions of "consciousness," "personal unconscious," and "collective unconscious" are assumed to be interrelated and to make decisions by processing some data.

The procedure for constructing a model of consciousness is as follows: First, we clarify the functions of conscious in Jungian psychology. Next, we consider the connections among the functions of consciousness, and the roles these connections play in each other. We define the functions of consciousness as shown in Table 1.

Table 1: Functions of Consciousness

Main components	Functions
Consciousness	Able to retrieve information from the unconscious that is related to the recognized information.
	Able to elicit needs and desires
	Able to repress needs and desires
Personal Unconscious	Complex (a desire or want that has been strongly suppressed by experience, education, and so on.)
	Suppressed needs and desires (currently unconscious needs and desires)
Collective unconscious	Human instincts (maternal instincts, sexual desires, and other archetypes shared by all human beings)

Assuming that the consciousness, personal unconscious, and collective unconscious have functions as shown in Table 1, there should be the interaction of data transfer and processing among main components.

Therefore, as a result of considering the data transfer and processing between the functions of consciousness as a decision-making system, Figure 1 was obtained.

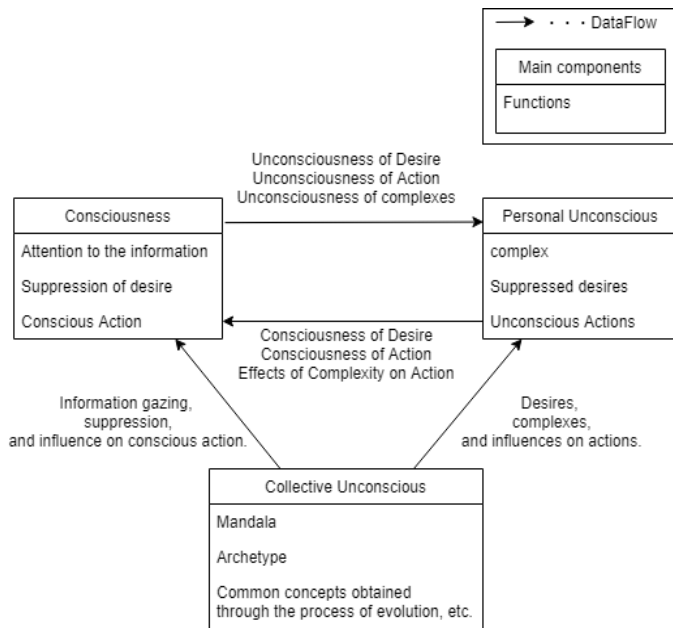


Figure 1: Conceptual diagram of consciousness

Figure 1 shows the connections between consciousness, the personal unconscious, and the collective unconscious, but this is too abstract to be reproduced as a program on a computer.

There must be connections between processes in the consciousness mind, connections between processes in the unconscious mind, and connections between small processes between the conscious mind and the unconscious mind.

Therefore, we further extended Figure 1 to consider the connections between each process and created Figure 2.

3.4. Overview of the Consciousness Model

By extending the conceptual diagram of consciousness in Figure 1, we obtained the following Figure 2. Using this figure as a model of consciousness, we can consider the operation of each process.

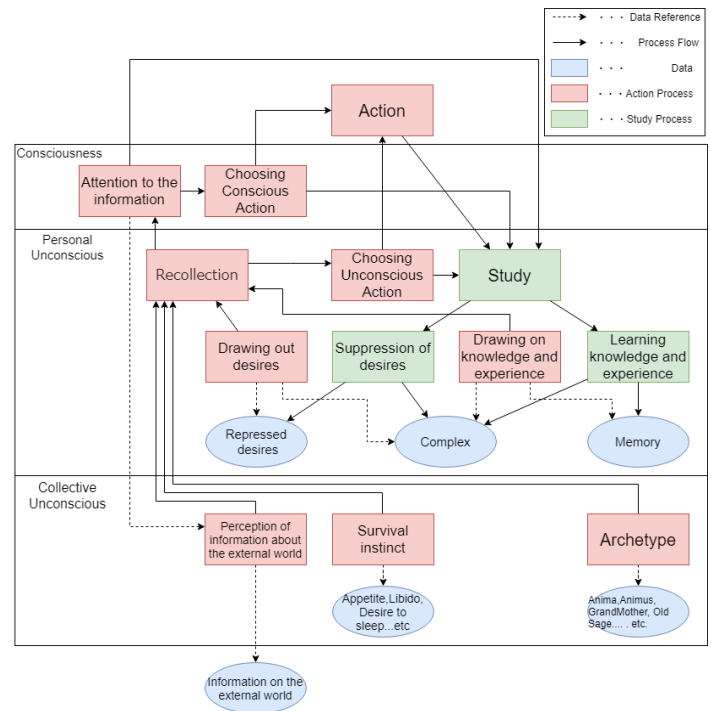


Figure 2: Consciousness Model

The process shown in Figure 2 can be divided into three major parts: "data," "Action process," and "Study process". For data, these include values obtained from sensors and episodes. The arrows between data and process represent data reference.

The action process is the process of taking action based on the data, and the study process is the process of retaining and storing the data as experiences and desires in the individual's memory after taking action.

For example, when a sensor detects hunger, the survival instinct perceives hunger and recalls associates information about food and how to eat it. Then, priming memory and other memories are sent to unconscious behavioral processes to select actions. Episodic memory and other memories are sent to information gazing to link the perception of external information with the information in episodic memory. After that, the conscious action selection process selects the action, and the action process selects the final action from the conscious and unconscious actions.

The focused information and selected behaviors are sent to the learning process, where the information is divided into the desire suppression process and knowledge-experience learning, and is prepared in the form of each memory before being stored.

The followings are the explanations of the components in Figure 2.

- Action

The actual action is taken by the system or the signal that directs it. The process of acting is selected by conscious and unconscious action selection.

- Information on the external world

Information that is not internal to the system, such as the five senses in humans, or sensors in machines. And it is unprocessed one, such as what you see with your camera or eyes, or what you feel.

- Attention to the information

To pay attention to objects and phenomena that are highly related to the objects recognized through the recognition of information about the external world, and to perform a deeper recall of these objects and phenomena.

- Choosing Conscious Action

To select the higher rewarding action required by the information you have been paying attention to.

- Choosing Unconscious Action.

To select highly learned behaviors picked up by recall.

- Recollection

The process of retrieving retained information in the right place and at the right time. Memory recall is the process of recalling a specific memory from among the memories stored in the brain. It induces memory destabilization, re-fixation, extinction learning, and memory association. In terms of the movement of the process, it refers to the action of searching for knowledge and experience to achieve desires.

- Drawing out desires

To search for and elicit relevant desires.

- Drawing on knowledge and experience

To search for and elicit relevant knowledge.

- Study

The act of filling in the blanks of knowledge. It is close to the meaning of the word in psychology and pedagogy and is distinguished from learning. It refers to the overall act of learning, such as extracting, categorizing, and storing features from the information obtained at that time as a result of an action, and adjusting each parameter. Knowledge is all that is gained through experience and includes all information stored in the unconscious. The rewards received from actions and related information are parameterized and recorded into knowledge, experiences, wants desires, and complexes.

- Suppression of desires

To repress what could not be done or failed in the drawn-out desire-lust.

- Learning knowledge and experience

To record the results and rewards of actions taken, and to store a series of actions as episodic memory.

- Repressed desires

A set of data repressed by desire.

- Memory

It is classified into sensory memory, short-term memory, and long-term memory, and it records information about what happens to us and what we feel, both temporarily and long-term. Sensory memory has the shortest storage time and is present in each sense. Only the information to which we pay attention becomes short-term memory. Short-term memory lasts only a few tens of seconds and is limited in the amount of information that can be stored at one time. Long-term memory is a part of short-term memory that is retained and becomes long-term memory through memory fixation. In this study, data is stored by learning, and organized by study.

- Perception of information about the external world

To recognize the type of object and distance from the external world information.

- Survival instinct

It elicits life-sustaining desire parameters such as sexual desire, appetite, and desire to sleep. The survival instinct is considered to be part of the collective unconscious.

It is the minimum means of survival stored as data.

- Archetype

To draws information about repression related to external world information and desires from stored data, such as anima and animus. Instincts and concepts cultivated in the group are individually shared data, not necessarily shared things in common.

Of the collective unconscious, the history and common sense cultivated by humankind fall into this category.

- Complex

In this study, we believe that complexes can be treated as a cohesive group of closely related memories and impulses, and as a collection of related memories. Complexes are said to be complexes of impulses, memories, and ideas that are intertwined by some emotion. By using these complexes, we will be able to handle memories that are highly relevant to recall. The parameters stored in the memory are treated as data that is associated with a distribution or set.

- Appetite, Libido, Desire to sleep...and so on.

Internal information such as a sensation of human internal organs, mental state, drowsiness, and so on., and remaining battery power and continuous operation time of the robot.

- Anima, Animus, Great Mother, Old Wise Man...and so on.

Conceptual things that humanity has in common, such as a sense of fear.

4. Discussion and further works

4.1. Discussion

The assumed environment and data are built on the premise that the input is received from the sensors installed in the computer, and the action is derived from the recall from this model.

In the future, when we implement this consciousness model, we need to consider various methods for each process.

Therefore, to clarify the data flow of this model and to make it easier to implement, we created a DFD(Data Flow Diagram) that is supposed to be installed in a robot or a simulator, and as a result, we obtained the following Figure 3.

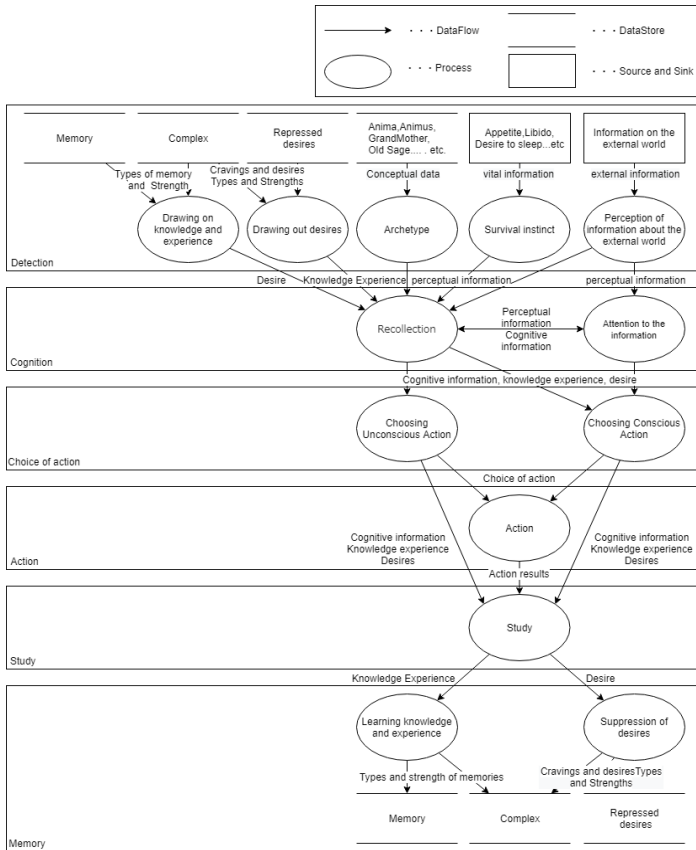


Figure 3: Diagram of consciousness DFD

Figure 3 shows the DFD version of the consciousness model in Figure 2. This DFD indicates that data will be processed from the top and passed to the next process. The arrows represent the direction in which the data is passed, and the names between the arrows represent the data.

Figure 3 shows that the flow of data is roughly in the following order: detection, data retrieval, action selection, action, study, and memory. This is roughly in line with the order of the process of detection-cognition-judgment-action, which is called the basic human thought process (<https://ipsj-is.jp/isdic/1070/>). We believe that this model can reproduce the basic thought process of human beings and still meet the requirements of the consciousness model defined in this study.

To implement the model on the agent in Unity for model validation, we will create a class diagram from Figure 2 for object-oriented implementation. Based on the obtained diagram, we plan to code and implement the model with the help of other researchers.

4.2. Further works

It will be necessary to implement the consciousness model in computers and robots to prove whether the model can function as

a decision-making program and make decisions based on an understanding of human ethics and values.

When we implement the consciousness model in the future, we need to be especially careful about the methods we use for each process, because some methods may cause unintended behavior. For example, the data structure of memory should be based on the structure of memory used in psychology and brain science, and the implementation should be based on human evidence.

In the learning process, it is also necessary to consider specific implementation methods on computers, such as the learning method and how to process the data to be handled.

5. Conclusion

In this study, we have developed a model of consciousness as a decision-making system based on Jungian psychology for implementation on computers and robots.

In the future, various issues remain to be solved in implementing this model, but we believe that solving them one by one will not only solve the ethical problems of AI but also lead to a better understanding of human thinking and brain mechanisms.

References

- [1] T. Watanabe, H. Kameda, "Proposal of a computer-reproducible model of consciousness," in 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS), IEEE: 1-4, 2020, doi:10.1109/SCISIS50064.2020.9322763.
- [2] M. Watanabe, K. Ogawa, H. Ishiguro, "Minami-chan: Application and Verification of Androids to Society Through Selling," *57*(4), 1251-1261, 2016.
- [3] T. Harada, S. Yamamoto, K. Yui, H. Kanda, Y. Chiba, H. Kameda, "Healing Robot PDDIN - Design and Implementation of prototype system -," The 65th National Convention Lecture and Paper Series, **2003**(1), 71-72.
- [4] H. Myouki, Introduction to Freud, ChikumaShinsho, 2000.
- [5] I. Kishimi, Introduction to Adlerian Psychology: For Better Relationships, bestsellers, 1999.
- [6] H. Kawai, Introduction to Jungian Psychology, baifukan, 1967.
- [7] M. Oikawa, H. Oikawa, "Consciousness and the unconscious in self-regulation: The effects of conscious compilation on goal priming," The Japanese Journal of Psychology, **81**(5), 485-491, 2010, doi:10.4992/jjpsy.81.485.
- [8] H.K. Walker, W.D. Hall, J.W. Hurst, Clinical Methods, 3rd edition, Butterworth, 1990.
- [9] S. Dehaene, H. Lau, S. Kouider, "What is consciousness, and could machines have it?," Science, **358**(6362), 486-492, 2017, doi:10.1126/science.aan8871.
- [10] N. Okada, "A Computational Model of the Mind," IEICE ESS Fundamentals Review, **1**(4), 30-45, 2008, doi:10.1587/essfr.1.4_30.
- [11] T. Maeno, "How to Make a Conscious Robot-Fundamental Idea based on Passive Consciousness Model-," Journal of the Robotics Society of Japan, **23**(1), 51-62, 2005, doi:10.7210/jrsj.23.51.
- [12] H. Chie, H. Takato, N. Takayuki, "OS-18 Emotion and AI," Artificial Intelligence, **34**(6), 881-887, 2019, doi:10.11517/jjsai.34.6_881.