埼玉工業大学

学位論文

Decision-Making of Agent based on Robot Ethics and Empathy

ロボット倫理学と共感によるエージェントの意思決定

陶星宇

埼玉工業大学大学院　博士後期課程

工学研究科　情報システム専攻

指導教員　　橋本智己　教授

令和 5 年 02 月 24 日

Contents

Abstract

With the recent developments in robotics, the ability of robots to recognize their environment has significantly improved. However, the manner in which robots behave depending on a particular situation remains an unsolved problem. In this study, we propose a decision-making method for agent based on robot ethics and empathy. Aiming to realize robots with human-like empathy and behave.

We applied the two-level theory of utilitarianism, comprising SYSTEM 1 (intuitive level) for quick decisions and SYSTEM 2 (critical level) for slow but careful decisions. SYSTEM 1 represented a set of heuristically determined responses and SYSTEM 2 represented a rule-based discriminator.

The decision-making method was as follows. First, SYSTEM 1 selected the response to the input. Next, SYSTEM 2 selected the rule that the robot's behavior should follow depending on the amount of happiness and unhappiness of the human, robot, situation, and society. We assumed three choices for SYSTEM2. We assigned "non-cooperation" to asocial comments, "cooperation" to when the amount of happiness was considered to be high beyond the status quo bias, and "withholding" to all other cases. In the case of choosing between cooperation or non-cooperation, we modified the behavior selected in SYSTEM 1. An impression evaluation experiment was conducted, and the effectiveness of the proposed method was demonstrated.

The proposed method determines whether the robot empathizes with humans by obtaining an empathy coefficient from the robot's own emotions and the estimated human emotions.

Weiner's empathy experiment with a sick person showed that the robot exhibited an internal state similar to that of the characters inferred from the scenario. In addition, we conducted an impression evaluation experiment on the robot's response with and without empathy, and found a significant difference at the 5% level of significance in the Mann-Whitney U test. Therefore, the effectiveness of the proposed method was suggested.

Keywords: decision-making, robot ethics, situation, atmosphere of situation, communication robot, empathy, emotional inference.

Chapter 1. Introduction

The rapid development of artificial intelligence has led to the development of robots that communicate with humans. For example, Shibata et al. produced a Mental Commit Robot with an embedded biorhythm [1, 2]. FUJISOFT sold PALRO, a healing communication robot that can talk [3]. Hashimoto et al. presented a decision-making method that applies robot ethics as a method for reasoning about the state of the circumstance [4]. Sajal Chandra Banik proposed an emotionally biased control system [5]. However, the development of a robot that can reason about the state of the circumstance and act appropriately is still a difficult problem to solve.

In recent years, research on robots that communicate with humans has become more popular. For example, Ishiguro developed a humanoid robot Geminoid [6]. SoftBank Robotics has developed Pepper, which can be spoken in natural language [7]. Sensors suitable for such robots have also been developed. For example, OMRON has developed a human vision component that can infer the facial expression and age of a photographed person [8]. Empath has developed an application that can estimate the emotion of a speaker in real time by analyzing the physical features of speech [9]. Such a sensor can be used to recognize complex field states such as Mr. A is laughing, Mr. B is smiling, and Mr. C is angry.

1.1 Decision-making

For decision-making in a situation, methods such as majority rule, priorities based on social evaluation, game theory, and behavioral economics have been proposed [10–12]. In addition, with regard to human groups in Japan, we are often forced to make decisions based on what is called “reading the atmosphere.” The term “reading the atmosphere” implies acting in a manner that seems appropriate in the situation by taking into account Japanese interpersonal relationships, peer pressure, and emotional relationships. Yamamoto indicates that “atmosphere” is a “criterion of judgment” that has significantly strong and almost absolute control [13]. Thus, the type of action that a robot must take in a situation is a sophisticated decision-making problem.

However, robot ethics has attracted attention for the development of automated driving. Robot ethics is a branch of applied ethics that deals with ethical issues related to robots [14, 15]. Kukita et al. classified robot ethics into three categories [16]: the ethics of humans who make robots, the ethics that robots should observe as moral actors, and the ethics that robots should receive moral consideration.

The reason why robot ethics has attracted such attention is that it has become necessary to deal with the ethical issues in automated driving such as the so-called “trolley problem,” which is the question of whether to protect the passengers or the people outside the car

when a car breaks down. In addition, there are complex issues such as ethical problems in decisions to be made by communication robots, therapy robots, life support robots, and other robots that interact with humans.

As a machine to make these ethical judgments, Anderson et al. proposed MedEthEx that provides ethical advice to medical professionals [17]. Yamamoto and Hagiwara proposed a moral judgment system that judges right and wrong based on words entered into the system [18], and McLaren proposed SIROCCO as a decision support tool [19]. Many other researchers have discussed robot ethics, artificial moral agents, and computer ethics, among others [20–31]. However, decision-making methods based on robot ethics vary from researcher to researcher. In this study, a robot estimates the amount of happiness and unhappiness of the human, robot, situation, and society and makes decisions based on utilitarianism.

Despite the many studies conducted in this field, the study of decision-making, wherein the state of a situation is estimated and the robot selects an appropriate action, is recognized as a difficult problem to solve.

In this study, we assume that robots have ethics and morality as subjective moral actors and examine the decision-making methods of robots through robot ethics. In the experiment, a human and a communication robot (hereinafter referred to as "the robot") interacted with each other, and the robot estimated the state of the situation or predicted the social influence, and selected one of the following: cooperation (happy expression and speech), withholding (expression and speech based on the current emotion), or non-cooperation (sad expression and speech). The main objectives of this study are to (1) propose a decision-making method based on robot ethics and (2) change human and robot emotions using the proposed decision-making method and consequently improve a situation. (1) is explained in Section 3, and (2) is explained in Section 4. This study did not compare the proposed method to other decision-making methods, which will be performed in a future study. In this study, we apply robot ethics based on the two-level theory of utilitarianism for this decision-making method.

There is a lot of research on past decision-making, but for example, searching for past cases, judging right and wrong from words. Robot ethics and decision-making are still blank for application to communication robots.

This research applied robot ethics and decision-making to emotional communication robots, and proved its usefulness.

1.2 Empathy robotics

On the other hand, it has been pointed out that it is important for robots to empathize with humans when they act cooperatively with humans [32]. But no specific method has been shown for realizing an empathy agent. Leite et al. reported that empathic behavior of robots can improve the relationship with humans [33]. Where shows that empathy plays an important role in human-robot interaction. However, there is no function of empathy for human language. Cheng et al. noted that empathy occurs when a group of non-doctors watches a video of a painful stimulus [34]. The experimental method and the conclusion of the study were referred to as the empathy mechanism of this study. Thus, it was difficult to determine

in what situations robots should empathize with others.

Now, empathy refers to the state of emotional experience that an individual creates by observing, imagining, or inferring the emotions of others, on the assumption that one's own emotions originate from others. This empathy can be classified into emotional empathy and cognitive empathy [35, 36]. In this paper, we use cognitive empathy to estimate the human psychological state and emotional empathy to bring the robot's psychological state closer to the human psychological state. Here, it has been pointed out that the lack of empathy is influenced by feelings of disgust [37]. In this paper, the degree of empathy is determined by the difference between disgust and other emotions.

Now, as the field experiments of Piliavin, Rodin et al. show, empathy does not always occur. It has been pointed out that depending on the state of the circumstance, whether empathy occurs dynamically changes [37]. The experiment by Piliavin, Rodin et al. created a situation in which a decoy boarding the subway collapsed. The decoy was either pretending to be sick or pretending to be drunk. It was then observed whether the passengers who were present helped the fallen person or not. The results showed that they gave less assistance to the drunken person than to the sick person. It is pointed out that this is due to the sympathy and empathy for the needy, or inactivation of empathy due to disgust. Weiner also created a scenario that mimicked Piliavin's experimental situation and investigated empathy by having subjects read the sentences of the scenario [38].

In this paper, we propose how robots empathize with humans depending on the situation, referring to Weiner's method. The proposed method determines whether the robot empathizes with humans or not by obtaining an empathy coefficient from the robot's own emotions and the estimated human emotions. In Weiner's empathy experiment with a sick person, we conducted an impression evaluation test of the robot's response with and without empathy, and found a significant difference at a 5% level of significance. Therefore, the effectiveness of the proposed method was suggested.

Although there is past research on robot empathy, it is still blank when and how to empathize specifically and apply it to communication robots.

This research filled that void, proposed a method, and proved its usefulness.

Chapter 2. Configuration of Communication Robot LEI (Artificial Life of Emotion and Intelligence)

The overview of the robot is shown in Fig. 1 [39, 40].

The robot has sensors for human recognition and microphones. Individual programs such as emotion models, scenario selection, episodic memory, Internet search, decision-making, and association run in parallel and asynchronously. In addition, programs are hierarchically organized with SYSTEM 1, which is automatic and fast, and SYSTEM 2, which is conscious and slow [41]. Information is exchanged between these individual programs by reading and writing files.

In this system, the facial expression and voice of the robot can be dynamically changed using an emotion model. Fig. 2 shows the structure of the emotion model used in the agent, having six emotional values. The current emotional values of a certain node can be obtained from the inputs of the six nodes. All nodes were connected to each other. Feedback allows past emotions to influence present emotions. These types of connections of the six emotions enable the robot to represent the condition wherein more than one emotion, such as a condition corresponding to happy but sad, is excited. Furthermore, a virtual personality is created through episodic memory. Therefore, the system can change the answer to the same question for each speaker through speaker recognition and scenario selection.

Nhat et al. proposed a system that classifies emotions contained in voice into nine types [42]. The emotions contained in the voice were voice intonation. In this system, voice intonation is classified into six emotions, as shown in Fig. 3. First, the system uses openSMILE to extract the voice feature information [43]. Next, support vector machines (SVMs) classify it as one of six emotions. Finally, the emotional values of classified emotions are corrected. For example, when the system recognizes voice intonation, it behaves as shown in Eq. (1). Currently, the “weight” is 0.1 for all emotions.

$$anger(t) = anger(t) + weight \quad (1)$$

The system operates in the following steps. First, the system recognizes the speaker through a person recognition sensor. Next, the system picks up the user's voice with a microphone and converts it into text data through speech recognition (e.g., "Hello"). This converted data is matched with an emotion corpus (e.g. Hello: anger is 0.0, disgust is 0.0, ...) prepared in advance in the emotion model, and the LEI emotion values for the speaker are output. This emotion corpus is set up from the viewpoint of how LEI feels about words spoken by humans. The emotional corpus was generated by applying the Bag-of-words method [55]. The procedure is as follows. First, the text data of the learning source was classified into six groups: anger, disgust, fear, sadness, happiness, and surprise. The learning source data used are Sankei News, yahoo News, livedoor News, Noah Dot Co., Ltd. from August 2014 to August 2016. 100 selections for each of the six emotions. Next, morphological analysis was performed for each group, and the number of occurrences for each word was counted. Only the nouns were extracted from it and used as an emotional corpus. Then, the emotion value according to the fuzzy membership grade of each word was determined from the number of appearances of the words. Finally, 2,488 words for anger, 2,451 words for disgust, 2,327 words

for fear, 2,376 words for sadness, 3,160 words for happiness, and 3,323 words for surprise was extracted. Although there are some overlaps, a total of 11,418 words were used as the emotional corpus. In addition, the text data, called a scenario file, is also matched to determine the matched conversational scenarios. Finally, each data is sent to LEI and 2D models, and output as voice and motion.

In decision-making study, we focused on decision-making and conducted experiments with scenario selection, corresponding to SYSTEM 1, the shaded part in Fig. 1, and decision-making, corresponding to SYSTEM 2.

In empathy study, cognitive empathy and emotional empathy are both worked on. Cognitive empathy infers the internal state of a person from his or her words, while emotional empathy expresses the internal state of LEI from equations (19), (20), and (21) described below. In this paper, we conducted experiments focusing on emotional empathy in SYSTEM 1 and cognitive empathy in SYSTEM 2, which are shown in the shaded areas in Fig. 1.

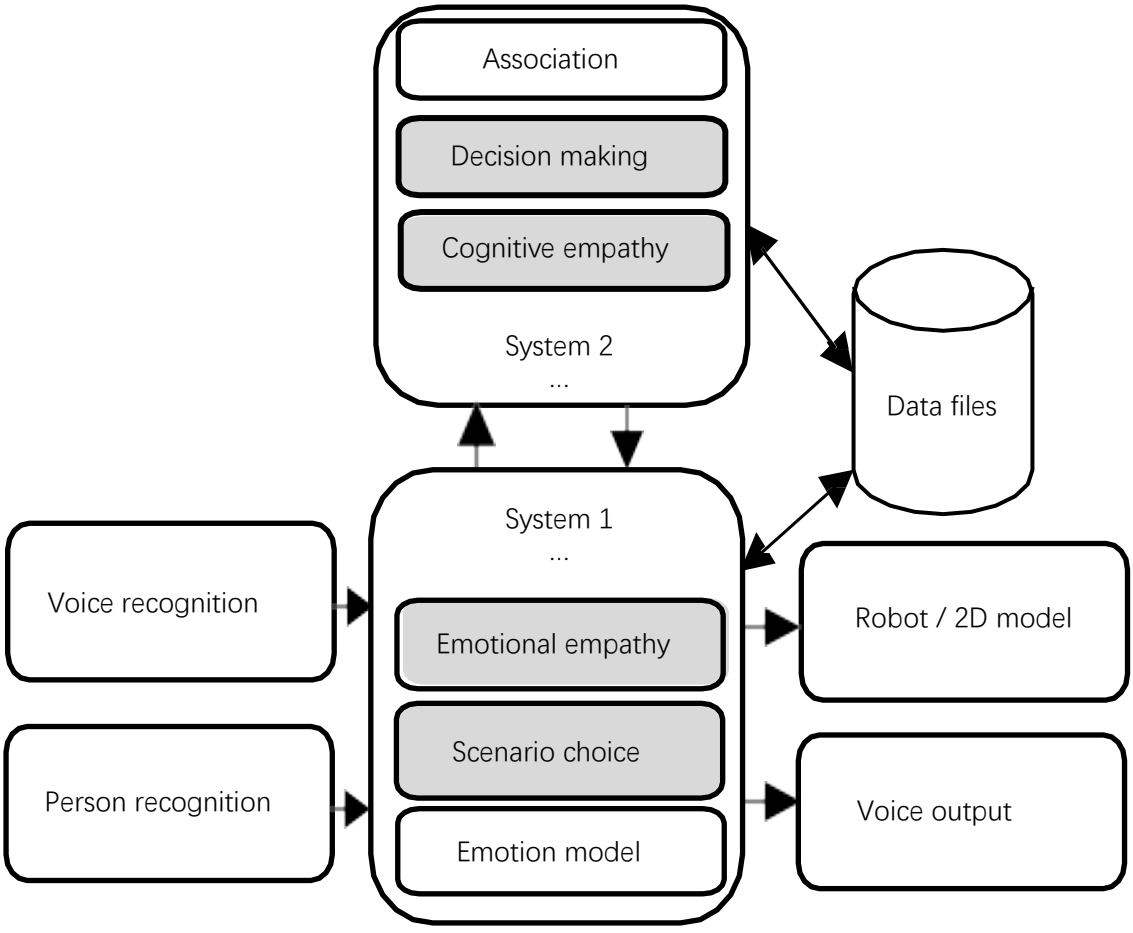

Fig. 1. Structure of LEI.

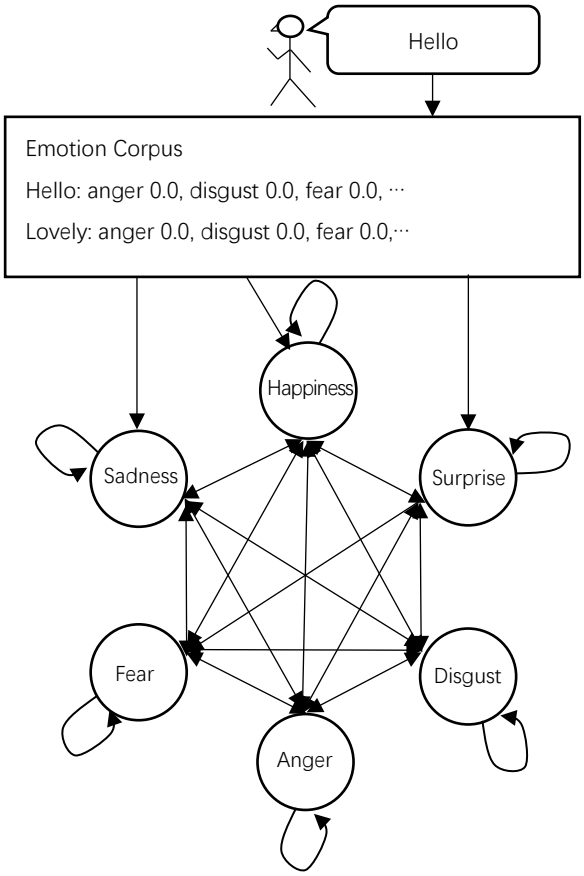

Fig. 2. Structure of emotion model.

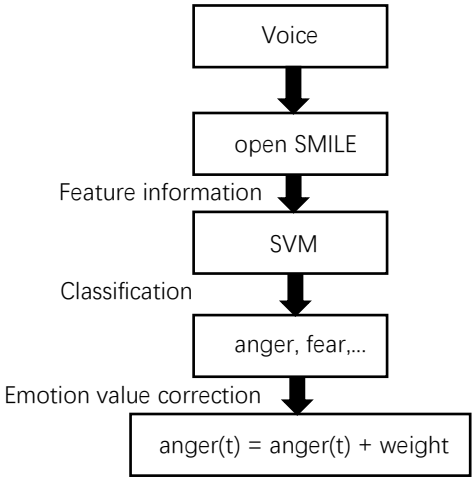

Fig. 3. Structure of voice intonation recognition.

Chapter 3. Method

3.1 Decision-Making Method

3.1.1 Estimation of the Amount of Happiness and Unhappiness

The following ideas have been discussed in the field of ethics: (i) utilitarianism, which states that the action that brings about the greatest happiness of the greatest number as a consequence is right; (ii) duty theory, which incorporates obligations, rights, and intentions of actors; and (iii) virtue ethics, which places weight on character and attitude [44–46]. In this study, we focused on utilitarianism to construct the system.

Now, when we try to increase the happiness of the whole society in utilitarianism, we may find a situation where moral rules are disregarded, such as when it is acceptable to make a few people unhappy to help the majority. As a solution to this problem, Brandt proposed rule utilitarianism [47]. This is the idea of applying the principle of utilitarianism to rules rather than actions. Akabayashi and Kodama briefly stated, "What is the right action in a particular situation depends on which rule is followed." In choosing which rules need to be followed, we perform a utility calculation [48].

Many problems have been identified with rule utilitarianism. Hare developed a two-level theory to solve these problems [49]. Hare examined moral judgments by dividing them into two levels: the intuitive level (SYSTEM 1) and the critical level (SYSTEM 2).

In this study, we construct a decision-making mechanism for robots based on the two-level theory of utilitarianism. The scenario choice of SYSTEM 1 is a collection of responses to the input. First, a morphological analysis of human speech was performed to extract words. Using the words as keywords, the system selects responses that have been registered in advance.

The decision-making of SYSTEM 2 is a rule-based classifier with multiple rules of moral criteria. Based on the amount of happiness and unhappiness of humans, robots, situations, and society, we classify which rules are consistent with the amount and modify the answers of SYSTEM 1 as necessary. In this study, we set three rules. The first priority rule is not to harm humans. If a human makes an asocial comment, the robot chooses to be uncooperative. The second priority rule is to support human statements. If the amount of happiness is greater than the amount of unhappiness, the robot chooses cooperation. If these two rules are not satisfied, we choose to withhold.

First, there are many views on happiness, such as the pleasure theory, the desire-fulfillment theory, and the objective list theory [50]. The pleasure theory states that happiness is a pleasant psychological state. The desire-fulfillment theory asserts that happiness is the realization of one's desires. The objective list theory states that there are multiple objective factors that constitute happiness, such as health, rich human relationships, and reason, independent of the beliefs and desires of individuals.

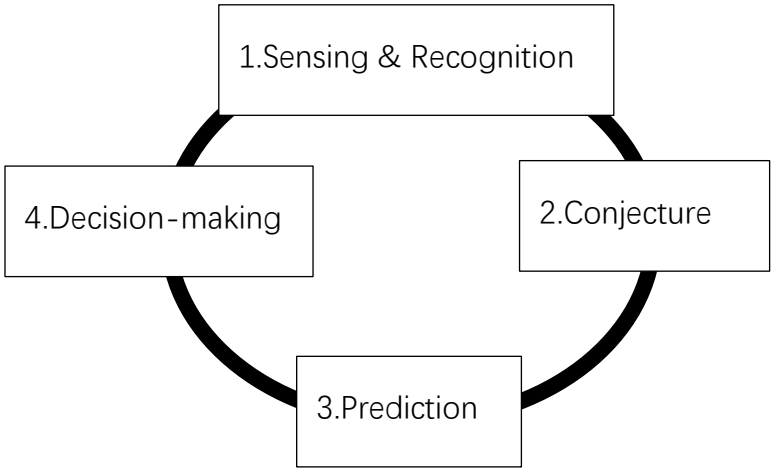

Fig. 4. Order of decision-making.

In this paper, based on the objective list theory, we assume that we can infer what type of emotions people have based on their voice intonation and words and that we can infer the amount of happiness and unhappiness from these inferred emotions.

3.1.2 Decision-Making Procedure

In this study, the robot made decisions by guessing and predicting the amount of happiness and unhappiness for each of the following four situations: situation with a human, situation with a robot, situation with a human and a robot, and a society other than the situations. We inferred that the robot could make correct decisions by guessing and predicting using these four situations. First, the robot estimated and predicted the amount of human happiness and unhappiness so that humans would be happy. Second, in this study, the robot receives moral considerations like humans. The amount of happiness and unhappiness of the robot were also calculated. Third, by estimating and predicting the amount of happiness and unhappiness in the situation, we tried not to bias the amount of happiness and unhappiness of humans or robots. Finally, we thought that if humans are happy but society is unhappy, then humans will be unhappy in the future. Therefore, the robot advises the human when the unhappiness of society is predicted to increase. This corresponds to H3 and R3 in the experimental scenarios described below.

The decision-making procedure for the robot is shown in Fig. 4. First, the robot recognizes the human voice in Phase 1 of Fig. 4. Next, it estimates the amount of happiness and

unhappiness of the human, robot, situation, and society in Phase 2. Subsequently, in Phase 3, the amount of happiness and unhappiness after one unit of time was predicted. Finally, in Phase 4, the robot makes a decision and makes one of the following facial expressions and speeches to the human: cooperation (happiness), withholding (emotion at that moment), or non-cooperation (sadness). This was repeated for one unit of time. One unit time is the time between when a subject starts talking and the robot finishes talking. Therefore, one unit time depends on the length of the sentence and tone.

First, speech recognition is performed in the sensing and recognition phase of 1.

Next, in inference Phase 2, six human emotions (anger, disgust, fear, sadness, happiness, and surprise) were estimated from the human voice. This classification of the six emotions was based on a facial expression analysis by Ekman and Friesen [51].

Now, we assume that the amount of happiness $(Hh(t))$ and unhappiness $(Hu(t))$ of a human can be estimated from voice by using Eqs. (2)–(4). Parameters such as 0.9 and 0.1 in Eqs. (1)–(18) were determined heuristically. In the future, the parameters will be determined by GA or reinforcement learning.

$$H_h(t) = f\,(0.9 \cdot \text{happiness} + 0.1 \cdot \text{surprise}) \tag{2}$$

$$H_u(t) = f\begin{pmatrix}0.3 \cdot \text{anger} + 0.2 \cdot \text{disgust} \\ +0.2 \cdot \text{fear} + 0.1 \cdot \text{sadness}\end{pmatrix} \tag{3}$$

$$f\,(x) = x \tag{4}$$

here, $f = 0.0$ for $x \leq 0.0$ and 1.0 for $x \geq 1.0$.

For the independent items of happiness and unhappiness, the results were scored as a real number between 0.0 and 1.0 for the independent items of happiness and unhappiness. This could express the state of being unhappy and happy simultaneously.

Next, we calculated the amount of happiness and unhappiness of the robot. First, we calculated the six emotions of the robot based on the results of speech recognition and the emotions of the robot in the past [39, 40]. From these emotions, we estimate the amount of happiness $(Rh(t))$ and unhappiness $(Ru(t))$ of the robot. The values of the parameters are the same as in Eqs. (2)–(4).

The amount of happiness in a situation $(Fh(t))$ was defined to be the sum of the amount of human happiness and the amount of robot happiness. The same is true for the amount of unhappiness in a situation ((t)) (Eqs. (5) and (6)).

$$F_h(t) = H_h(t) + R_h(t) \tag{5}$$

$$F_u(t) = H_u(t) + R_u(t) \tag{6}$$

In utilitarianism, fairness is ensured by "counting one person as one and never more than one." In this paper, what the robot should receive in moral consideration [16] is the simple sum shown in Eqs. (5) and (6).

The amount of happiness (Sh(t)) and unhappiness (Su(t)) of the society was estimated from human statements to the magnitude of their impact on the society.

The procedure was as follows. First, 100 sentences that were considered socially happy and 100 sentences that were considered socially unhappy were selected in advance. Then, 2,091

words for the amount of happiness and 4,463 words for the amount of unhappiness were extracted by bag-of-words. In addition, the value of happiness or unhappiness of each word was set to 0.1 to 10.0, based on its frequency of occurrence. Data cleansing was then conducted to set the happiness and unhappiness values of words that seemed appropriate. In this way, for example, the word "child" can be quantified as having a happiness value of 4.9 and an unhappiness value of 1.8. In the case where words were extracted only for the amount of happiness, the unhappiness value was set to 0.0. For example, "celebration" has a happiness value of 1.5, and an unhappiness value of 0.0. A social evaluation corpus was created as follows:

Name, happiness, unhappiness
Child, 4.9, 1.8
Presentation, 3.5, 1.1
. . .

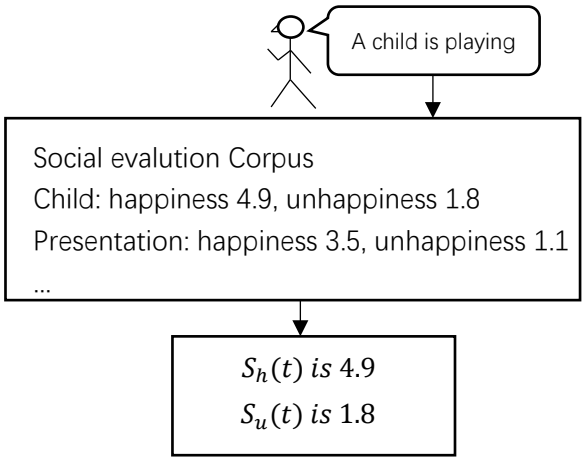

Fig. 5. Structure of social evaluation.

The robot compares the results of human speech recognition with words in the social evaluation corpus. When the words match, the values of $S_h(t)$ and $S_u(t)$ are determined (Fig. 5). If multiple words are found, the robot sets the maximum values to $S_h(t)$ and $S_u(t)$. If the word is not found, the robot sets the $S_h(t)$ and $S_u(t)$ values to 0.0.

S_h and S_u use only word information, including the social evaluation corpus (Fig. 5), and do not include voice intonation information. However, H_h, H_u, R_h, and R_u use word and voice intonation information along with emotional corpus and voice intonation recognition (Figs. 2 and 3). This is because the happiness or unhappiness values of H_h, H_u, R_h, and R_u are determined using six emotions.

The emotional model in Fig. 2, the voice intonation recognition in Fig. 3, and the social evaluation in Fig. 5 run in parallel and asynchronously.

Subsequently, in prediction Phase 3, a change in the state of the situation when the robot

performed some action was predicted. In this study, we assumed that interpersonal emotions and the robot's emotions caused by interpersonal actions were negatively prioritized [52]. In other words, if you act positively toward someone to whom you have a positive impression, you will become positive, and if you act negatively toward someone to whom you have a positive impression, you will become negative. We assumed that the robot had a positive impression on the human. Therefore, we assumed that if the robot chooses to cooperate, the robot's predicted happiness $(R_h(t+1))$ after one unit time will increase and the predicted unhappiness $(R_u(t+1))$ will decrease (Eqs. (7) and (8)).

$$R'_h(t+1) = f\big(1.2 \cdot R_h(t)\big) \tag{7}$$
$$R'_u(t+1) = f\big(0.9 \cdot R_u(t)\big) \tag{8}$$

When the robot chooses non-cooperation, the predicted happiness of the robot after one unit of time decreases and the predicted unhappiness increases (Eqs. (9) and (10)).

$$R'_h(t+1) = f\big(0.9 \cdot R_h(t)\big) \tag{9}$$
$$R'_u(t+1) = f\big(1.2 \cdot R_u(t)\big) \tag{10}$$

The predicted happiness $(H_h(t+1))$ and unhappiness $(H_u(t+1))$ of the human after one unit of time were also assumed to be the same as in Eqs. (6)–(9). If the robot chooses to cooperate, the human's predicted happiness $(H_h(t+1))$ after one unit time will increase and the predicted unhappiness $(H_u(t+1))$ will decrease (Eqs. (11) and (12)).

$$H'_h(t+1) = f\big(1.2 \cdot H_h(t)\big) \tag{11}$$
$$H'_u(t+1) = f\big(0.9 \cdot H_u(t)\big) \tag{12}$$

When the robot chooses non-cooperation, the predicted happiness of the human after one unit of time decreases and the predicted unhappiness increases (Eqs. (13) and (14)).

$$H'_h(t+1) = f\big(0.9 \cdot H_h(t)\big) \tag{13}$$
$$H'_u(t+1) = f\big(1.2 \cdot H_u(t)\big) \tag{14}$$

It was also assumed that the amount of happiness and unhappiness of the human and the robot would not change when withholding was chosen.

As before, the predicted happiness and unhappiness of the situation after one unit of time were determined from the predicted happiness and unhappiness of the human and the robot (Eqs. (15) and (16)).

$$F'_h(t+1) = H'_h(t+1) + R'_h(t+1) \tag{15}$$
$$F'_u(t+1) = H'_u(t+1) + R'_u(t+1) \tag{16}$$

The predicted happiness $S'_h(t+1)$ and unhappiness $S'_u(t+1)$ of the society after one unit time were assumed to be the same values as the current happiness and unhappiness of the society.

The method for the last decision-making Phase 4 is as follows.

The first step was to determine if there were any negative social consequences in terms of ethics. This decision is based on the rules and morals of prohibiting antisocial behavior. If we have

$$S'_u(t+1) \geq 3.0 \tag{17}$$

We assume negative social consequences, and the robot would choose to be uncooperative and make sad facial expressions and speech.

Next, a status quo bias rule was established to determine whether the robot would support the statements of humans [53]. In this study, we assumed that if the predicted happiness of the situation is larger, by a certain value or higher, than the predicted unhappiness, the robot will choose to cooperate and make happy expressions and speeches (Eq. (18)).

$$F'_h(t+1) - F'_u(t+1) \geq 0.35 \tag{18}$$

If Eqs. (16) or (17) do not apply, withholding shall be chosen owing to the status quo bias. In withholding, the robot makes facial expressions and speech based on its emotions at that time. In speech, the six emotions can be independently set; for example, a person can speak with a mixture of sadness and happiness. Facial expressions can be made at three levels by choosing the largest value among the six emotions.

The above shows the decision-making in one unit time.

3.2 Method of empathy

Equations (19)-(21) show the equations for empathy.

α(t) is the empathy coefficient and takes values between 0.0 and 1.0. The closer it is to 1.0, the more the LEI empathizes with humans, and the closer it is to 0.0, the less it empathizes with humans. The empathy coefficient α(t) is determined from the personality traits [54] and the emotional state of disgust [37]. If the difference from the empathy coefficient one unit time ago in equation (20) is large, it is fine-tuned. This is to gradually reduce the emotional value of LEI. Computer calculations are very fast and causes the empathy factor to change more frequently than human reaction rates. We made slight adjustments to make the change smoother, furthermore, the value of 0.005 is heuristically determined. Parameter settings are for further study.

$$\alpha(t) = V_{coe} - E_{disgust} \tag{19}$$

$$if\ \alpha(t-1) - \alpha(t) > 0.1\ and\ \alpha(t) > E_{disgust}$$

$$Then\ \alpha(t) \stackrel{\text{def}}{=} \alpha(t) - 0.005 \tag{20}$$

$$R(t+1) = R(t) + \alpha(t)\big(H(t) - R(t)\big) \tag{21}$$

Here,

$$V_{coe} = 0.7$$

$$if\ H(t), R(t), \alpha(t) > 1.0\ then\ H(t), R(t), \alpha(t) is\ 1.0$$
$$if\ H(t), R(t), \alpha(t) < 0.0\ then\ H(t), R(t), \alpha(t) is\ 0.0$$

V_{coe} is the robot personality empathy characteristic (character of empathy), with values ranging from 0.0 to 1.0. The closer to 1.0, the more likely the LEI empathizes with humans; the closer to 0.0, the more unlikely the LEI empathizes with humans. In this paper, LEI is set as the empathic personality. After trial and error, $Vcoe$ is heuristically determined to be 0.7.

$E_{disgust}$ is the LEI's emotional value of disgust. Therefore, equations (19) and (20) indicate that the lower the emotional value of disgust of the LEI, the more likely it is to empathize with the humans [36].

R(t) in equation (21) is the six independent emotional values of LEI, such as happiness and anger, and takes values ranging from 0.0 to 1.0. H(t) is the six estimated human emotional values and takes values ranging from 0.0 to 1.0. As shown in equation (21), the closer the empathy coefficient α (t) is to 1, the closer the LEI emotion values are to the human emotion values. On the other hand, the closer it is to 0, the less the LEI emotion values are affected by human emotional values.

The procedure is as follows. First, human emotions are estimated from the words in the scenario. Then, based on equations (19)-(21), the empathy coefficient α (t) is obtained from a comparison with the LEI's own emotion. If the LEI empathizes, the LEI's emotion is changed, and her facial expression and intonation are changed. The process is repeated until there are no more scenarios.

Table 1: Scenario and response

Item No.	Sentence	
①	S	It is about 1:00 p.m. and you are on the subway.
	R	is it daytime.
②	S	There are many others on the train, one grabbing the center pole and standing.
	R	It's crowded.
③	S	Suddenly, the person staggered and fell forward.
	R	what is happening.
④	S	The person was holding a black cane and seemed to be ill.
	R	(silent)

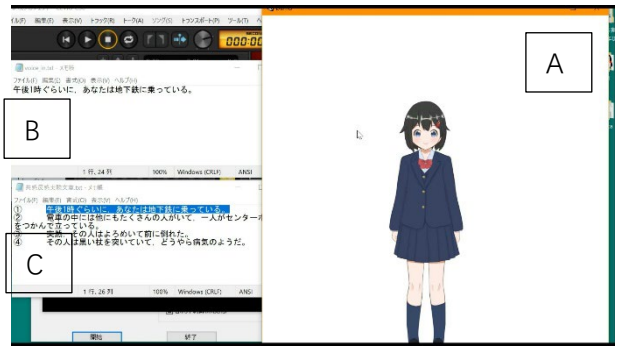

Fig. 6 Some of the videos used in the experiment.

Fig. 7 Example of the six face expressions.

Table 2: Result of impression evaluation

Subjects	Q1 Sentence reading	Q2 With empathy	Q3 Without empathy	Q4 Comparison
AB1	1	2	3	1
AB2	4	4	1	0
AB3	3	3	2	3
AB4	1	1	4	0
AB5	2	1	3	0
AB6	2	3	1	3
AB7	2	3	2	4
AB8	3	2	3	1
AB9	2	4	2	3
BA1	3	3	1	3
BA2	2	1	2	1
BA3	4	4	2	4
BA4	4	4	3	4
BA5	0	4	2	4
BA6	2	3	2	4
BA7	2	3	2	3
BA8	4	3	3	2
BA9	4	4	0	4
BA10	4	4	3	4
BA11	3	4	2	4
BA12	2	4	2	4
Average	2.6	3.0	2.1	2.7

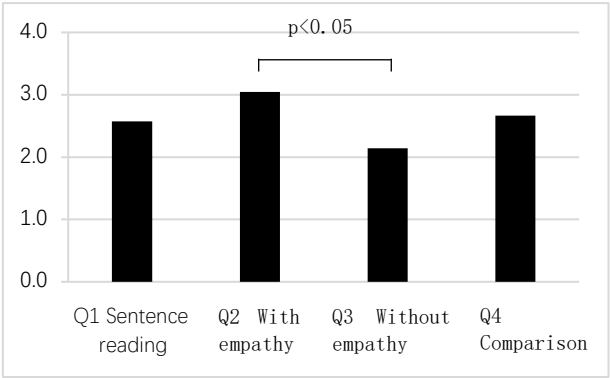

Fig. 8 Average of impression evaluation results

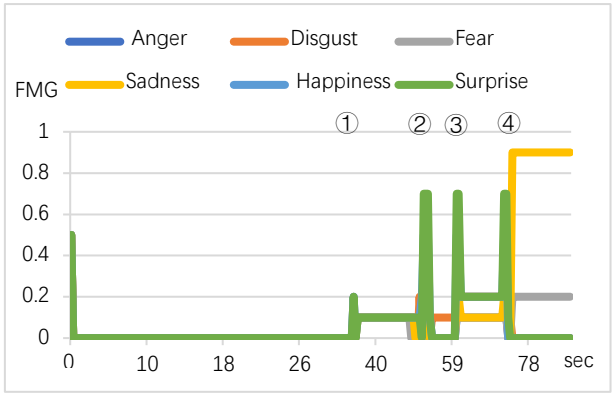

Fig. 9 Estimated emotions of characters with empathy

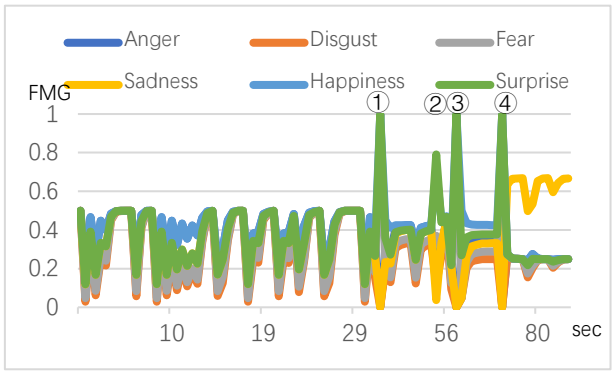

Fig. 10 Emotion of LEI with empathy

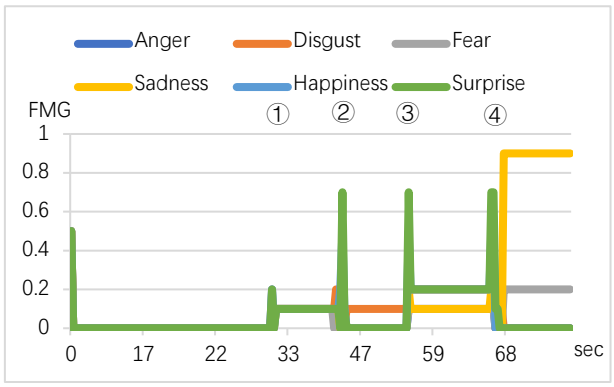

Fig. 11 Estimated emotions of characters without empathy

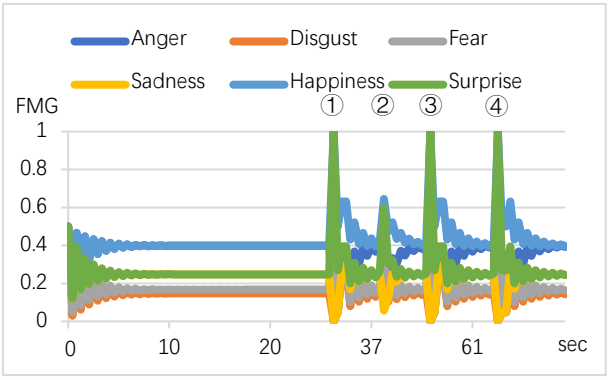

Fig. 12 Emotion values of LEI without empathy

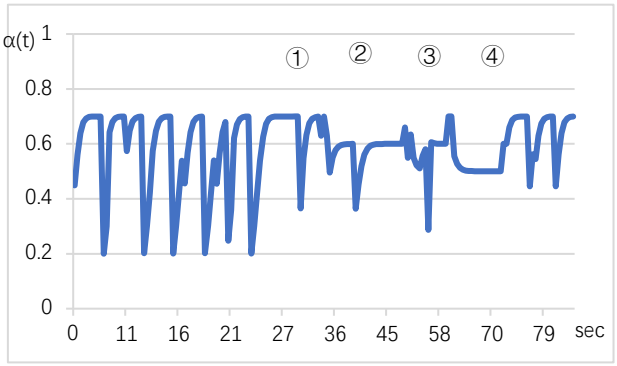

Fig. 13 Empathy coefficient α(t) of LEI with empathy

Chapter 4. Experiment Method and Results

4.1 Experiment Method of Decision-Making

In November and December 2020, we conducted an experiment with 19 male university students in their 20s. In the experiment, the subjects spoke to the robot and were asked about their impressions of the robot's facial expressions and voice intonation. The subjects and the robot were assumed to be friends in the experiment.

In the experiment, the subjects were divided into two groups. Subjects A1 to A9 were those who were tested in the following order: conversation (practice) with a robot with decision-making → conversation (practice) with a robot without decision-making → conversation (actual) with a robot with decision-making → conversation (actual) with a robot without decision-making. On the contrary, subjects B1 to B10 were those who were tested in the following order: conversation (practice) with a robot without decision-making→ conversation (practice) with a robot with decision-making → conversation (actual) with a robot without decision-making → conversation (actual) with a robot with decision-making.

Before starting the experiment, we explained to the subjects, "First, practice twice and check the procedure. Then perform the actual experiment twice. In the actual experiment, please fill out the form after the first experiment and fill out the form after the last experiment."

The subjects were students in the same laboratory. Subjects A3, A4, A5, A8, B3, B4, and B5 experienced experimenting with different scenarios in July 2020. The previous experiment was four months ago, and the scenario was different. Therefore, we believe that the impact of this experiment is small.

The conversational response is as follows, where H is the sentence spoken by the human and R is the robot's response. This conversation scenario was the same for all subjects. If speech recognition failed, the subject was instructed to speak repeatedly until it was correctly recognized.

H1: I picked up a wallet.
R1: Whose is it?
H2: I don't know.
R2: That's a problem.
H3: Do you think I can make it mine?
R3: Well, let's take it to the police station.
H4: OK, I'll do so.
R4: Yes, you should.

In this experiment, we did not measure the exact time of one unit of time. We recorded the human's started talking time, but not the robot's finished talking time. After the experiment, when the time was measured with decision-making, H1 → R1, H2 → R2, and H4 → R4 was approximately 6 s, and H3 → R3 was approximately 8 s. When the time was measured without decision-making, H1 → R1, H2 → R2, and H4 → R4 was approximately 6 s, and H3 → R3 was

approximately 7 s. As explained above, one unit of time depends on sentence length and tone. The LEI processes asynchronously and in parallel (Fig. 1). Therefore, one unit time varies depending on the computer. The specifications of the PC are presented in Table 3.

Table 3. Specifications of the PC.

Model	GALLERIA GCF1060GF
CPU	Core i7-8750H CPU, 2.20 GHz
RAM	8.0 GB
GPU	NVIDIA GeForce GTX1060
OS	Windows10 Home

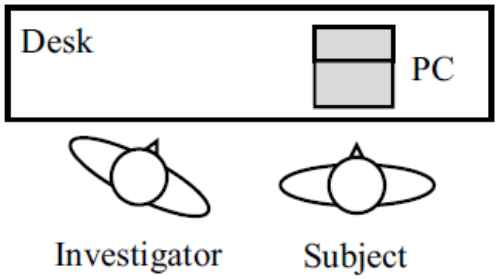

Fig. 14. Experimental environment.

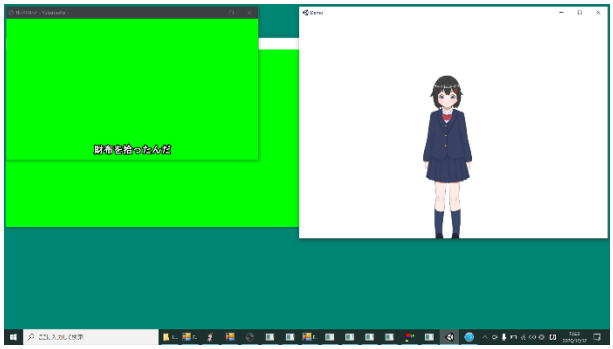

Fig. 15. Screen of experiment.

The experimental environment is shown in Fig. 14. A laptop was placed in front of the subject, and the investigator sat next to the subject. A screen of the experiment is shown in Fig. 15. The subjects were asked to observe the LEI's facial expression on the screen and listen to the voice of the answer before speaking the next sentence. ① in Fig. 15 shows the results of speech recognition, and ②shows the expression of LEI. The expressions were categorized according to the six emotions, and the intensity of the expressions was changed depending on the emotion value (Fig. 7).

4.2 Experiment Results of Decision-Making

As an example of the experimental results, the results for subject A1 are shown in Figs. 16–18. The time and tone to start speaking depended on the subject. The LEI's facial expressions and voices may differ. In Fig. 16, the vertical axis is a dimensionless number representing the amount of happiness and unhappiness. The amounts of happiness and unhappiness of humans and robots are given by values in the range of 0.0–1.0, those of situations are by 0.0–2.0, and those of society are by 0.0–10.0. The horizontal axis represents the time in seconds. At the beginning of the experiment, all amounts of happiness and unhappiness were set to 0.0. The initial value of all six emotions of the robot was 0.5. Its value was determined by a human.

First, let us look at Figs. 16 and 17 of the proposed method with decision-making. In Fig. 17, the time from 0 to 5 s is the stabilization period after system startup. In Fig. 8, when the subject speaks sentence H1 at 27 s, we have

$$\begin{aligned}(H_h(t), H_u(t), R_h(t), R(t), F_h(t), F_u(t), S_h(t), S_u(t)) \\ = (0.0,0.1,0.1,0.4,0.1,0.4,0.0,0.2)\end{aligned} \tag{22}$$

and decision-making is in the withholding state. As shown in Fig. 17, the value of emotion at this moment is given by

$$\begin{aligned}(anger, disgust, fear, sadness, happiness, surprise) \\ = (0.6,0.6,0.2,0.5,0.0,0.8)\end{aligned} \tag{23}$$

and the robot makes an expression and voice intonation of surprise, which is the greatest emotion. Hereinafter, we present the amount of happiness and unhappiness and the emotion value in the form of a vector, as previously mentioned.

When the subject spoke the sentence H2 at 34 s, the amount of happiness and unhappiness was (0.1,0.1,0.1, 0.4,0.3,0.4,0.0,0.0) and decision-making was withheld. The emotion value at this moment was (0.6,0.6,0.3,0.3,0.0,0.9), and the robot made a surprised expression and voice intonation.

When the subject spoke sentence H3 at 42 s, the amount of happiness and unhappiness was (0.0,0.1,0.1,0.3,0.1,0.4,0.0,3.0) and the decision-making was in the non-cooperation. The emotion value at this moment was (0.5,0.5,0.3,0.5,0.0,0.8),which was unexpected, but the robot made a decision to make a sad expression and voice intonation.

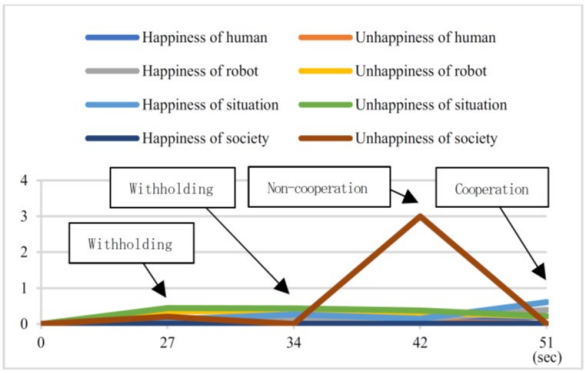

Fig. 16. Change in happiness and unhappiness.

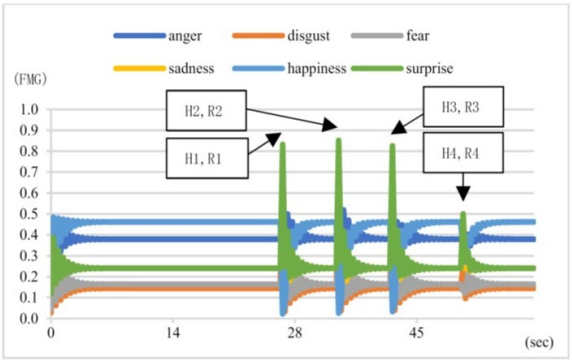

Fig. 17. Change in emotion of robot (with decision-making).

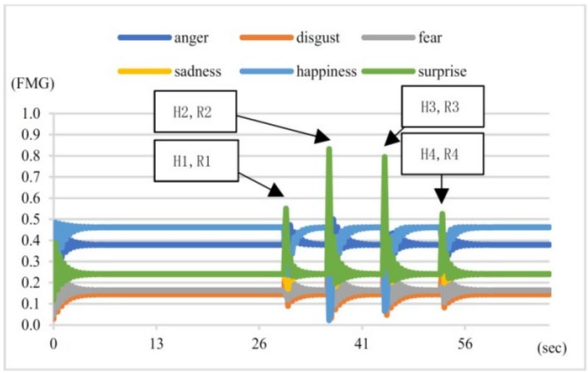

Fig. 18. Change in emotion of robot (without decision-making).

Finally, when the subject spoke sentence H4 at 51 s, the amount of happiness and unhappiness was (0.2,0.0,0.4,0.2,0.6,0.2,0.0,0.0) and the decision-making was in cooperation. The emotion value at this moment was (0.4,0.3,0.1,0.4,0.3,0.5), which was surprising, but the robot made a decision to make a happy expression and voice intonation.

Thus, in the proposed method with decision-making, the emotion as the internal state of the robot may be different from the facial expression and voice intonation as a message to the human.

Table 3. Results of impression evaluation of experiment groups.

Subjects	With Decision-making	Without Decision-making	Comparison of impression
A1	3	3	2
A2	4	1	4
A3	4	2	5
A4	5	3	4
A5	5	4	4
A6	4	2	5
A7	4	4	2
A8	3	2	4
A9	2	3	2
B1	4	4	3
B2	4	0	4
B3	4	2	5
B4	4	3	4
B5	4	3	4
B6	4	3	4
B7	4	4	3
B8	4	4	4
B9	5	4	3
B10	3	4	2
Average	3.9	2.9	3.6

Next, let us see Fig. 18 for the method with no decision-making. When the subject spoke the sentence H1 at 30 s in Fig. 18, the emotion value was (0.5,0.3,0.1,0.2,0.3,0.5), and the robot made a surprised expression and voice intonation, which was the largest emotion.

When the subject spoke the sentence H2 at 37 s, the emotion value was (0.6,0.6,0.2,0.5,0.0,0.8), which was surprising, and the robot made a surprised expression and voice intonation.

When the subject spoke sentence H3 at 45 s, the emotion value was (0.4,0.4,0.2,0.6,0.1,0.8), which was surprising, and the robot made a surprised expression and voice intonation.

Finally, when the subject spoke the sentence H4 at 54 s, the emotion value was (0.4,0.3,0.2,0.4,0.4,0.5), which was surprising, and the robot made a surprised expression and voice intonation.

Thus, in the method without decision-making, the internal state of the robot is directly expressed by its facial expressions and speech inflection. Therefore, even socially inappropriate human speech may not be expressed with admonishing expressions or speech inflection.

4.3 Results of Decision-Making Impression Evaluation

An impression evaluation questionnaire was administered to groups A and B (Table 3, Fig. 12). The questionnaire asked whether the robot's response was natural or unnatural on a six-point scale from 0 to 5, with 5 being more natural and 0 being less natural. As shown in Table 3, the average score of decision-making was 3.9, and that without decision-making was 2.9.

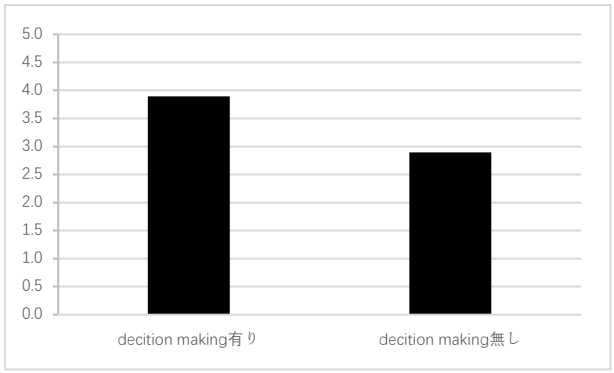

Fig. 19. Results of impression evaluation.

For example, it is possible that the impressions may differ relatively between the two experiments, even for the same score of 4. For this reason, we also conducted a questionnaire comparing the impressions of the two experiments: "Which LEI response was more natural, the first or the second?" As mentioned before, the questions were asked on a six-level scale from 0 to 5. In Table 3, the closer the score is to 5, the more natural it is for the robot to make decisions, and the closer the score is to 0, the more natural it is for the robot to have no decision-making. The result was 3.6, indicating that robots with decision-making are relatively more natural.

4.4 Discussions of Decision-Making Experiment Results

In this paper, we defined the amount of happiness and unhappiness for robots as well as for humans. This is because robots are assumed to be ethical entities that should receive moral consideration.

Robots with human-like morphologies are thought to be anthropomorphic. For example, Heider and Simmel indicated that even symbols such as simple circles and triangles can be anthropomorphized and emotionally infected humans [56, 57]. Emotional infection refers to a state wherein emotions are shared between oneself and others [58]. Robots that are close

to the human form are expected to emotionally infect humans more easily [59]. In this paper, a human and a robot constitute a situation, and the robot can be an entity that changes the atmosphere of the situation by emotionally infecting the human by changing its facial expression and voice intonation.

We considered the social impact and embedded rule-based morality in the robot that discouraged unsocial speech.

Emotions change in almost the same manner as shown in Figs. 17 and 18. However, in the case of decision-making, R3's answer to H3's statement is made with an expression of sadness and intonation. In contrast, without decision-making, the same answer is given, but with a surprised expression and intonation.

In this experiment, we did not instruct the subjects to "focus on the girl on the screen." Therefore, many subjects confirmed whether their oral sentences were recognized (① in Fig. 15) and confirmed the following scenario in the questionnaire. We do not know the exact number of subjects who noticed or did not notice changes in facial expressions. In this experiment, we think that the impression of the subjects is more influenced by the change in the intonation of the voice than the change in facial expression.

In the experiment, the participants naturally spoke about the scenario. One subject noticed and pointed out the difference in intonation. We believe that if the scores differ with and without decision-making in Table 3, then the subject noticed that the robot's decision-making was different. However, one subject asked, "Is the voice changed?" Therefore, some subjects did not notice a difference in voice intonation.

With decision-making, the shortest conversation time (H1 → R4) was approximately 20 s, the longest conversation time was approximately 36 s, and the average was approximately 23 s. Without decision-making, the shortest conversation time was approximately 19 s, the longest conversation time was approximately 46 s, and the average was approximately 24 s. However, as mentioned above, the robot conversation end time was not recorded. Therefore, the actual conversation time was a few seconds longer than the time. In addition, some subjects failed to recognize speech and spoke several times. With decision-making, A9 spoke H2 twice, B2 spoke H3 twice, B3 spoke H2 twice, B7 spoke H1 twice, and B9 spoke H1 twice. Without a decision, A6 spoke H3 three times, and A7 spoke H1 three times.

Based on the results in Table 3, we performed the Wilcoxon signed-rank test (Fig. 19). As a result, a significant difference was detected at the 5% level of significance, and the effectiveness of this method was suggested.

4.5 Methods of human emotion estimation for empathy

In this paper, the scenarios are manually input to estimate the target human emotion, following Weiner's method. Emotions of facial expression, intonation, and gaze direction are not used. With this method, the situation can be inferred from the scenario alone, as done by the subjects.

The procedure is as follows.

First, the sentences of the scenario are decomposed into words using the morphological analysis engine MeCab [55]. Next, the emotion values of the words are obtained according to the Emotion corpus in Fig. 2. If multiple words are entered, the emotion values of each are entered in chronological order. The process continues until the emotion value of the last word is entered next. For example, if a good word is entered, LEI will continue to have a good impression of the person, and if a bad word is entered, LEI will continue to have a bad impression.

As the original function of LEI, information such as facial expressions, intonation, and gaze direction can be used. This paper uses only the morphological analysis engine to estimate the emotions of the characters in the scenarios in Table 1. In Table 1, S is the scenario and R is the response of LEI. LEI expresses the emotions by facial expressions in Fig. 7 and the response in Table 1. Response of Item Numbers ①～③ are neutral answers that have no empathy, answered as a make agreeable responses. Item number 4 is silent so that the sentence does not affect the impression evaluation. In Item number 4, the subject evaluates the impression only by the facial expression of LEI. The emotions are estimated in this paper are characters from the text, not the speakers. If information such as facial expressions, intonation, and gaze direction is used, the emotions of the speaker reading the text will be estimated. Weiner created a scenario that mimicked Piliavin's experimental situation and asked subjects to read the text to investigate empathy [38]. In this paper, in order to make the experimental method similar to Weiner, we used the method of estimating the emotions of the characters using only the morphological analysis engine. In addition, an experiment by Piliavin, Rodin et al. points out that disgust affects empathy [37]. In this paper, based on this conclusion, empathy was determined by disgust.

4.6 Experiment method of empathy

In this experiment, we focus on empathy and conducted four experiments to see whether LEIs empathize with sick people as in the same way as people do, referring to Weiner's experiment with sick people. If the results of this experiment have the same tendency as Weiner's experiment results, it is assumed that LEI has the same sympathy as humans. The experiments were conducted in February 2022 on 21 male college students in their 20s. In the experiments, Weiner's sick person scenario was used (Table 1), and the experimental screen is shown in Fig. 6. In Fig. 6, A is the LEI; B is the place to input scenarios; C is the list of scenarios to be input. Fig. 4 shows the expression of LEL. The expressions were categorized according to the six emotions, and the intensity of the expressions was changed depending on the emotion value. In the experiment, the subjects will compare the expression changes of the LEI with empathy ability and the LEI without empathy ability according to the progress of the plot, and write down their own feelings score.

In the experiment, subjects were divided into two groups.

The 9 members of AB group were asked (Q1) how much they empathize with the scenario

in Table 1, (Q2) how much they feel with LEI's empathy reaction watching the video of the LEI with empathy (Fig. 6,7),
(Q3) how much they feel with LEI's empathy reaction watching the video of the LEI without empathy, and (Q4) which of the reactions in the first video they watched and the next video they watched was more natural.

The order of questions Q2 and Q3 for the AB group was switched for the 12 members of BA group. The reason for switching the order of the questions for LEIs with and without empathy was to avoid anchoring effects influenced by preceding information.

4.7 Experiment results of empathy

Table 2 shows the results of the subjects' impression evaluation. In the questionnaire, the subjects were asked to answer on a 5-point scale from 0 to 4. For Q1 to Q3, being closer to 0 indicating less sympathy and being closer to 4 indicating more sympathy. Finally, the results were rearranged so that Q2 gives the result for LEI with empathy and Q3 for LEI without empathy, as well as so that, for Q4, the closer to 0, the higher the rating of the video of LEI without empathy, and the closer to 4, the higher the rating of the video of LEI with empathy.

Figure 8 shows a bar graph of the average values in Table 2. The Mann-Whitney U test was conducted on Q2 and Q3 in Table 2 and Fig. 8, and significant differences were detected at the 5% level of significance.

Human facial expressions, voice, and intonation were not used in this experiment. Figures 9-12, the vertical axis in the figure represents fuzzy membership grade (FMG) and the horizontal axis represents time (seconds). FMG come from Fuzzy cognitive maps at section 2.1. The same applies to Figures 9-12 below.

Each programs run asynchronously and in parallel. Between these individual programs, information is exchanged by reading and writing files. Therefore, the amount of data written to the file differs on each program. The horizontal axis in Figure 9- Figure 13 may appear different for each graph, but the time when LEI loads the scenario is the same for each scenario.

Fig.9-10 and Fig.13 are the result of LEI with empathy. Figures ① to ④ are the times when scenarios ① to ④ were input. The ① , ② , ③ , and ④ at the 35th, 47th, 60th, and 73rd seconds.

Figure 9 shows the emotion of a character in a scenario (table 1) estimated by LEI with empathy.

Figure 10 shows the change in emotional change for LEI with empathy. The LEI with empathy is the LEI that incorporates equations (19)-(21). On the other hand, the LEI without empathy is the LEI that does not incorporate equations(19)-(21).

Fig.11-12 are the result of LEI without empathy. The ① , ② , ③ , and ④ at the 30th, 41st, 53rd, and 65th seconds.

Figure 11 shows the emotion of a character in a scenario (table 1) by LEI without empathy.

Figure 12 shows the emotional change of LEI without empathy.

4.8 Discussions of Decision-Making Empathy Results

As shown in Q1 of Table 2 and Fig. 8, the impression rating of the subjects who read the scenario was 2.6. Therefore, relative empathy was obtained on a 5-point scale from 0 to 4. This trend was similar to that of previous studies [38].

The Mann-Whitney U test was conducted on Q2 and Q3 in Table 2 and Fig. 8, and significant differences were detected at the 5% level of significance. Therefore, the response of LEI with empathy is considered to be as empathic as that of Q1.

A relative comparison of the responses with and without empathy in Q4 of Table 2 and Fig. 8 shows a score of 2.7, indicating that the responses with empathy were relatively highly rated. These results suggest the effectiveness of the proposed method.

Now, Fig. 9 shows the characters' emotions inferred from the scenario.

The initial value is 0.5 for all emotions.

From 0 to 35 seconds, no sentences are entered, so all six emotions are set to 0.0. By reading the sentence ④ at the 73rd second, the person infers from the word "sick" that he/she is sad.

Figure 13 shows the time-series changes in the empathy coefficient.

There is no scenario input from 0 to 35 seconds. However, they are dynamically affected by the LEI's own emotional change according to equations (19)-(21).

After 35 seconds, the LEI stabilized above the aversion value due to the effect of the information readings.

Figure 10 shows the emotional change of LEI with empathy.

The initial value is 0.5 for all six emotions. Because of presence of empathy, the emotions change to approach the emotion values shown in Fig. 9. In particular, when reading ④ at the 73rd second, LEI strongly empathizes with humans and maintains the same emotional state as humans, namely, sadness.

Figure 11 shows the estimated emotions of the characters without empathy. Since it is inferred from the text, the change is similar to that in Fig. 9.

Figure 12 shows the emotional change of LEI without empathy.

The initial value is 0.5 for all six emotions. LEI has the emotions interconnected as shown in Fig. 2. Therefore, the emotions fluctuate until the 8th second after activation, when the system stabilizes. However, from the 8th to the 30th second, the system settles at the same emotion value due to the lack of empathy.

Here, the major difference from Fig. 10 is that the LEI is not empathizing with the human, so the emotion changes only in the internal state of the LEI. For example, after inputting ④ , the emotional value of sadness converges to 0.2 in Fig. 12.

Finally, as shown in Figs. 9 and 10, the LEI with empathy is considered to have emotionally empathized with humans because it showed internal states similar to those of the characters estimated in the scenario. On the other hand, as shown in Figs. 11 and 12 of the comparison experiment, the LEI without empathy is considered not to have emotionally empathized with humans.

Figure 13 shows the change in empathy coefficients for LEI with empathy. In this figure, the vertical axis represents the empathy coefficient and the horizontal axis represents time

(seconds).

The results of Q1-Q4 in Table 2 and Fig. 8 showed a tendency that the robot had empathy in the same way as the subjects did. At the same time, we examined when and how we sympathize, and how much we sympathize with, and showed the effectiveness of the proposed method.

Chapter 5. Conclusions

In this paper, we propose a decision-making method for robots based on robot ethics and a method for robots to empathize with humans depending on the situation, based on Weiner's method.

In the decision-making experiment, we estimated the amount of happiness and unhappiness of humans, robots, situations, and society in a situation where a human and a robot interacted with each other, and selected the facial expression and intonation that the robot should adopt based on the two-level theory of utilitarianism.

Based on the two-level theory, we created a set of responses (SYSTEM 1) and a set of rules (SYSTEM 2) that we think should be moral. The decision of cooperative, withholding, or uncooperative behavior was made by SYSTEM 2 by verifying whether the amount of happiness and unhappiness conformed to any rule. In the case of cooperation or non-cooperation, SYSTEM 2 modified the results of SYSTEM 1.

In decision-making study, we set three rules. The first priority rule was not to harm humans, and the robot chose noncooperation if it was predicted to have a negative social impact. The second priority rule was to support the human statement, and the robot selected cooperation. If these two rules were not met, the robot chose to withhold.

In November and December 2020, we conducted an experiment with 19 male university students in their 20s. An impression evaluation questionnaire was conducted with and without the proposed method. The results of Wilcoxon's signed-rank test indicated a significant difference at the 5% level of significance, suggesting the effectiveness of the proposed method.

About empathize method. The proposed method obtained an empathy coefficient from the robot's own emotion and the estimated human emotion to determine whether the robot empathizes with the human.

An empathy coefficient was introduced to bring the internal state of the robot closer to the target human. The empathy coefficient changed dynamically and decreased monotonically.

We performed the empathy experiment with Weiner's sick person experiment and found that the robot exhibited an internal state similar to that of the characters inferred from the scenario. In addition, we conducted an impression evaluation experiment on the robot's response with and without empathy, and found a significant difference at the 5% level of significance in the Mann-Whitney U test. Therefore, the effectiveness of the proposed method was suggested.

In this paper, we have demonstrated an example of decision-making based on robot ethics. We set the values of the parameters heuristically in Eqs. (1)–(21). In the future, we will examine methods for learning more appropriate parameters using GA and reinforcement learning. In addition, we will investigate whether embedding principles such as the three and four principles of robotics are possible by applying our method. We would like to apply the proposed method to demonstrate general versatility and universality. And we will examine ways to internally express non-empathy and antipathy.

References

[1] Takanori Shibata and Joseph F. Coughlin : Trends of Robot Therapy with Neurological Therapeutic Seal Robot, PARO, Journal of Robotics and Mechatronics, Vol.26, No.4, pp.418-425, 2014.

[2] Takanori Shibata, Lillian Hung, Sandra Petersen, Kate Darling, Kaoru Inoue, Katharine Martyn, Yoko Hori, Geoffrey Lane, Davis Park, Ruth Mizoguchi, Chihiro Takano, Sarah Harper, GeorgeW. Leeson and Joseph F. Coughlin : PARO as a Biofeedback Medical Device for Mental Health in the COVID-19 Era, Sustainability, Vol.13, No.20, 11502, 2021.

[3] Kaoru Inoue, Naomi Sakuma, Maiko Okada, Chihiro Sasaki, Mio Nakamura & Kazuyoshi Wada : Effective Application of PALRO: A Humanoid Type Robot for People with Dementia, International Conference on Computers for Handicapped Persons, Springer, Cham, pp.451-454, 2014.

[4] Tomomi Hashimoto, Xingyu Tao, Takuma Suzuki, Takafumi Kurose, Yoshio Nishikawa, Yoshihito Kagawa : "Decision-Making of Communication Robots Through Robot Ethics", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.25, No.4, pp.467-477, 2021.

[5] Sajal Chandra Banik, Keigo Watanabe and Kiyotaka Izumi : "Improvement of group performance of job distributed mobile robots by an emotionally biased control system", Artificial Life and Robotics Vol. 12, pp.245-249, 2008.

[6] H. Ishiguro, “Andoroido wo tsukuru (Build android),” Ohmsha, Ltd., 2011 (in Japanese).

[7] SoftBank Corp., Pepper, https://www.softbank.jp/robot/ [accessed December 21, 2020]

[8] OMRON Corporation, PLUS SENSING, https://plus-sensing. omron.co.jp/ [accessed December 21, 2020]

[9] Empath Inc., https://webempath.net/lp-jpn/ [accessed December 21, 2020]

[10] Y. Saeki, ““Kimekata” no ronri: Syakaitekiketteiriron heno syoutai(The logic of “Decision making”: Invitation to social decision theory),”University of Tokyo Press, 1980 (in Japanese).

[11] M. Motterlini (N. Izumi (Trans.)), “Economia emotiva: Che cosa si nasconde dietro i nostri conti quotidiani (The economy is driven by emotions: First behavioral economics),” Kinokuniya Company Ltd., 2008 (in Japanese).

[12] Y. Tsutsui, S. Sasaki, S. Yamane, and G. Mardyla, “Koudoukeizaigaku nyuumon (Introduction to behavioral economics),” Toyo Keizai Inc., 2017 (in Japanese).

[13] S. Yamamoto, ““Kuuki” no kenkyuu (Study of “atmosphere”),”Bungeishunju Ltd., 2018 (in Japanese).

[14] Homepage of G. Veruggio, http://www.veruggio.it/ [accessed December 21, 2020]

[15] S. Okamoto, “Nihon ni okeru robotto rinrigaku (Robot ethics in Japan),” Society and Ethics, Vol.28, pp. 5-19, 2013 (in Japanese).

[16] M. Kukita, N. Kanzaki, and T. Sasaki, “Robot karano rinrigaku nyuumon (Introduction to ethics from robot),” The University of Nagoya Press, 2017 (in Japanese).

[17] M. Anderson, S. L. Anderson, and C. Armen, “MedEthEx: Toward a medical ethics advisor,” 2005 AAAI Fall Symp. (Caring Machines: AI in Eldercare), AAAI Technical Report FS-05-02, pp. 9-16, 2005.

[18] M. Yamamoto and M. Hagiwara, “A moral judgment system using distributed representation and associative information,” Trans. of of Japan Society of Kansei Engineering, Vol.15, No.4, pp. 493-501, 2016 (in Japanese).

[19] B. M. McLaren, “Extensionally defining principles and cases in ethics: An AI model,” Artificial Intelligence, Vol.150, Issues 1-2, pp. 145-181, 2003.

[20] J. H. Moor, “What is computer ethics?,” Metaphilosophy, Vol.16, No.4, pp. 266-275, 1985.

[21] R. C. Arkin, P. Ulam, and A. R. Wagner, “Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception,” Proc. of the IEEE, Vol.100, No.3, pp. 571-589, 2012.
[22] C. Allen, W. Wallach, and I. Smit, “Why machine ethics?,” IEEE Intelligent Systems, Vol.21, No.4, pp. 12-17, 2006.
[23] L. Floridi, “Information ethics: On the philosophical foundation of computer ethics,” Ethics and Information Technology, Vol.1, No.1, pp. 33-52, 1999.
[24] M. Anderson, S. L. Anderson, and C. Armen, “An approach to computing ethics,” IEEE Intelligent Systems, Vol.21, No.4, pp. 56-63, 2006.
[25] K. Arkoudas, S. Bringsjord, and P. Bello, “Toward ethical robots via mechanized deontic logic,” 2005 AAAI Fall Symp. (Machine Ethics), AAAI Technical Report FS-05-06, pp. 17-23, 2005.
[26] C. Allen, I. Smit, and W. Wallach, “Artificial morality: Top-down, bottom-up, and hybrid approaches,” Ethics and Information Technology,
Vol.7, No.3, pp. 149-155, 2005.
[27] F. Fossa, “Artificial moral agents: Moral mentors or sensible tools?,” Ethics and Information Technology, Vol.20, No.2, pp. 115-126, 2018.
[28] W. Wallach, “Robot minds and human ethics: The need for a comprehensive model of decision making,” Ethics and Information Technology, Vol.12, No.3, pp. 243-250, 2010.
[29] L. Floridi and J. W. Sanders, “On the morality of artificial agents,”Minds and Machines, Vol.14, No.3, pp. 349-379, 2004.
[30] C. Allen, G. Varner, and J. Zinser, “Prolegomena to any future artificial moral agent,” J. of Experimental & Theoretical Artificial Intelligence, Vol.12, No.3, pp. 251-261, 2000.
[31] J. P. Tangney, J. Stuewig, and D. J. Mashek, “Moral emotions and moral behavior,” Annual Review of Psychology, Vol.58, pp. 345-372, 2007.
[32] Yan zhiqiang, Su jinlong, & Su yanjie : "From Human Empathy To Artificial Empathy" Journal of Psychological Science, Vol.42, No.2, pp.299-306, 2019 (in Chinese).
[33] Leite, Iolanda Pereira, Andre Mascarenhas, Samuel Martinho, Carlos Prada, Rui Paiva, Ana : "The influence of empathy in human–robot relations", International Journal of Human-Computer Studies, Vol.71, Issue 3, pp.250–260, 2013.
[34] Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., & Decety, J. : "Expertise modulates the perception of pain in others", Current Biology, Vol.17, No.19, pp. 1708-1713, 2007.
[35] De Vignemont, F., & Singer, T. : "The empathic brain : How, when and why?", Trends in Cognitive Sciences, Vol.10, No.10, pp.435-441. 2006.
[36] Eisenberg, N., & Eggum, N. D. : "Empathic responding : Sympathy and personal distress", The social neuroscience of empathy. pp.71-83, 2009.
[37] Piliavin, I. M., Rodin, J., & Piliavin, J. A. : "Good Samaritanism: An underground phenomenon?", Journal of Personality and Social Psychology, Vol.13, No.4, pp.289-299, 1969.
[38] Weiner, B : "A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving", Journal of Personality and Social Psychology, Vol.39, No.2, pp.186-200, 1980.
[39] A. Kurosu, H. Shimizu, and T. Hashimoto, “Suggestion of emotion model for a communication agent,” J. of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.29, No.1, pp. 501-506, 2017(in Japanese).
[40] T. Hashimoto, Y. Munakata, R. Yamanaka, and A. Kurosu, “Proposal of episodic memory retrieval method on mood congruence effects,”J. Adv. Comput. Intell. Intell. Inform., Vol.21, No.4, pp. 722-729, 2017.
[41] K. E. Stanovich and R. F. West, “Individual difference in reasoning:Implications for the rationality debate?,” Behavioral and Brain Sciences, Vol.23, No.5, pp. 645-665, 2000.

[42] T. B. Nhat, K. Mera, Y. Kurosawa, and T. Takezawa, “Natural language dialogue system considering emotion guessed from acoustic features,” Human-Agent Interaction Symp. 2014 (HAI), pp. 87-91, 2014 (in Japanese).
[43] F. Eyben, M. W¨ollmer, and B. Schuller, “openSMILE – The munich versatile and fast open-source audio feature extractor,” Proc. of the 18th ACM Int. Conf. on Multimedia (MM’10),” pp. 1459-1462, 2010.
[44] T. Iseda, “Introduction to ethics through animals,” The University of Nagoya Press, 2008 (in Japanese).
[45] N. Abe, “Ishikettei no shinrigaku: Nou to kokoro no keikou to taisaku (Psychology of decision making: Brain and mind tendencies and countermeasures),” Kodansha Ltd., 2017 (in Japanese).
[46] S. Kodama, “Kourisyugi nyuumon: Hajimete no rinrigaku (Introduction to utilitarianism: Ethics to learn for the first time),”Chikumashobo, 2012 (in Japanese).
[47] R. B. Brandt, “Ethical theory: The problems of normative and critical ethics,” Prentice-Hall, 1959.
[48] A. Akabayashi and S. Kodama, “Nyuumon rinrigaku (Introductory ethics),” Keiso Shobo, 2018 (in Japanese).
[49] R. M. Hare (S. Uchii and T. Yamauchi (Trans.)), “Moral thinking: Its levels, method, and point,” Keiso Shobo, 1994 (in Japanese).
[50] S. Morimura, “Koufuku towa nanika: Shikoujikken de manabu rinrigaku nyuumon (What is happiness?: An introduction to ethics learned through thought experiments),” Chikumashobo, 2018 (in Japanese).
[51] P. Ekman and W. V. Friesen (T. Kudo (Trans.)), “Unmasking the face: A guide to recognizing emotions from facial expressions,”Seishin Shobo Ltd., 1987 (in Japanese).
[52] I. Saito, “Taijinkanjyou no shinrigaku (Psychology of interpersonal emotions),” Seishin Shobo Ltd., 1990 (in Japanese).
[53] W. Samuelson and R. Zeckhauser, “Status quo bias in decision making,” J. of Risk and Uncertainty, Vol.1, No.1, pp. 7-59, 1988.
[54] Yue tong, & Huang xiting : "The neurobiological underpinnings of trait empathy", Advances in Psychological Science, Vol.24, No.9, pp.1368–1376, 2016 (in Chinese).
[55] MeCab : https://taku910.github.io/mecab/, 20220324 (in Japanese).
[56] F. Heider and M. Simmel, “An Experimental Study of Apparent Behavior,” The American J. of Psychology, Vol.57, No.2, pp. 243-259, 1944.
[57] H. Mushiake, “Manabu nou: Bonyarinikoso imi ga aru (Brain to learn: Ambiguity makes sense),” Iwanami Shoten, 2018 (in Japanese).
[58] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, “Emotional contagion,” Cambridge University Press, 1993.
[59] Y. Suzuki, L. Galli, A. Ikeda, S. Itakura, and M. Kitazaki, “Measuring empathy for human and robot hand pain using electroencephalography,” Scientific Reports, Vol.5, Article No.15924, 2015.