human/non-human interface Workshop “After-Cyborg”
Graduate School of Social Sciences, Hitotsubashi University, Tokyo, Japan
3rd February, 2018

Toward the Age of the Social Cyborg

The Cyborganization of Human Society through Agent Technology

OSAWA, Hirotaka *

I named the talk “Toward the Age of the Social Cyborg: The Cyborganization of Human Society through Agent Technology,” which sounds a little confusing. So, I would like to briefly explain the content.

My name is Osawa and I am an assistant professor of system information at the University of Tsukuba. First, I would like to tell you about myself. Earlier, it was mentioned that I am a researcher in the field of engineering, but I will tell you what type of research I am doing. In short, I am researching human–agent interaction. I do not believe many of you are familiar with the term, so I would first like to explain what it means. To put it simply, it is research that deals extensively with interactions between humans and characters such as those operated by artificial intelligence.

Some researchers regard it as the broad study of a phenomenon called “agency attribution,” in which people find themselves feeling otherness. We often summarize the nature of our work as dealing with the anthropomorphization of artificial intelligence or social intelligence.

As for the content of the research, it is varied, but I would like to provide some concrete examples by showing videos, etc.

For example, this anthropomorphizes electrical appliances and makes them explain themselves regarding instructions on how to use them (Figure 1). Now, you may feel as though you just saw something strange and unexpected. That reaction is precisely what I study. I test the outcomes of things like creating a character for a machine to make it introduce itself.

* Faculty of Engineering, Information and Systems, University of Tsukuba.

This is a tool used to test the various expressions that can be created according to the user's personality, for instance, by strengthening or weakening the anthropomorphism.

Figure 1

We have discovered that if we use such characters, people will listen to the explanations more carefully than they would with other agents.

These multifunction printers have a robot-type agent attached to them, and when it explains something, it retracts inside the printer; that is to say, it makes it seem as though it has gone into the screen, and provides instructions there (Figure 2). This way, it provides a visual display that uses the advantages of both the space inside the screen, called “virtual space,” and the space outside the screen, making it appear as though the two are seamlessly connected. I have

Figure 2

studied the value of that. For example, I have found that it can increase your motivation or actually make the instructions easier to understand.

This is called a werewolf game. (Figure 3) It is a game that is played these days, particularly by young people. To put it simply, it is like a mafia game. In my research, I create agents who actually play this game and study the influence of the gamer’s emotional expressions, etc.

Figure 3

Here is someone who is wearing glasses, but there are eyes on these glasses, and they move, for example, according to your partner's eyes rather than your own eyes. (Figure 4) This research work might enable us to communicate by directing our gaze to the same place as our partner, while performing other tasks, such as sensing social moods. Now, I would like to tell you more about why I am doing this kind of research.

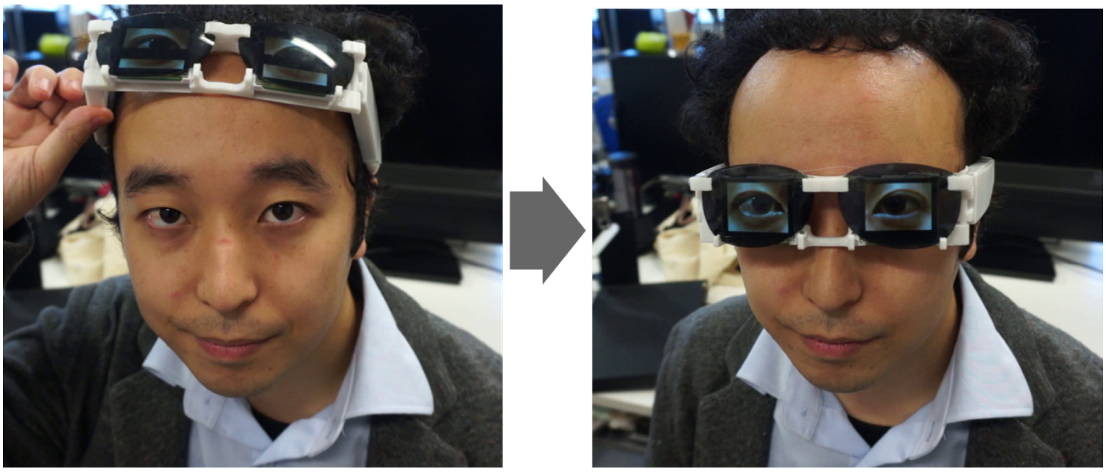

Figure 4

First, my personal background is computer-related. I have been studying computers for a long time, particularly in the Department of Information Engineering. As I was working there, I gradually began to study robots and interfaces, and I became particularly interested in anthropomorphism among robotics. That is how I ended up conducting this kind of research.

I think it would be easier to understand if I explained my personal history. Since I am of the generation that was born in the 1980s and grew up after video games had been created, I am a common example of someone progressing from playing those games to becoming interested in programming.

When I was in high school, I was interested in things like artificial intelligence and artificial life, and I read books about them. The 1980s were the time of a second AI boom, and various books about the subject came out in those years. As I was reading them, I found that AI was believed to be difficult to develop at the time, and I became aware that one of the problems was that it would not have an actual body as its environment. That sparked my interest in the development of AI with a body, and I went in that direction at university.

At university, I belonged to the so-called SF research group and Robot Technology Research Group, but I am not quite sure about whether I learned theories in the SF research group; however, Keio University's SF research group included members such as Takayuki Tatsumi and Mari Kotani. They are the people who translated Donna Haraway's book, and I feel that we are in relatively similar fields in that sense. Today, listening to that talk, I recalled some fond memories of those years.

At that point, I had decided that I wanted to do research related to artificial intelligence when I entered the laboratory during my fourth year at the university. After seeking out a myriad of information, I decided that I wanted to study artificial intelligence in the laboratory of Dr. Yuichiro Anzai, former president of Keio University. Although artificial intelligence was a difficult field and was not attracting a lot of interest at the time, I was advised to study properly, beginning with computer basics, and I later studied under Prof. Imai in the Anzai laboratory. At the time, I was quite interested in things like machine learning, but one of my advisors told me that those fields had already been researched quite a lot and that the most interesting field at the time was the one in which open systems interacted with humans. I did not understand it very well then, but later, when I actually joined the lab, I had an experience of meeting a humanoid robot called Robovie that was built to communicate, and it held me in its arms. (Figure 5)

Figure 5

Robovie was actually operated by a very simple algorithm. When the robot said, "hold me," and spread its arms, a sensor measured the distance between it and the user. Although it was a very simple machine that anyone could build, I was rather impressed by its performance. How can I describe it? It felt unexpected and surprising. It was a simple device, but it worked really well, so much so that I had the impression that it held me as a response to my holding it. I thought that this kind of field was interesting in the sense that something that I had never thought about or imagined was happening there. Thus, I entered the somewhat bizarre field of robots that communicate with people.

There are a lot of things that I thought about while studying robots. At first, I used the humanoid robot Robovie, but gradually, I began to think about various things like finding out what may be called "contradictions." For example, there were different plans at that time as to whether these robots could actually enter society or be placed inside a home. It was just before the Aichi Expo, and there were quite a lot of studies on robots, but while I was doing research, even though I heard many people saying, "In the future, robots will be like this...," I wondered whether that was really the case and whether it was really what was sought after.

I thought, "Hardware is not necessarily that important for the robots we are talking about here." When I thought about what it meant to have been moved by a robot giving me a hug, I came to realize that what is important about robots is not the hardware itself but rather how people feel in their experiences with robots. At that point, I knew what I should do is reproduce *that* part—the feelings people have when interacting with robots—and I began to study the anthropomorphism or agents, that is, what it means to make people feel what we call "otherness."

Today, I am doing a new type of research called human–agent interaction. This is a concept that various researchers have gradually begun to consider since the 2000s. Basically, the general definition given by Prof. Yamada—who is now the chairman of the Japanese Society for Artificial Intelligence—is that it is research that deals with autonomous systems that interact with the world outside humans and systems that give the impression that they are autonomous.

To put it a little more simply, I try to tell my students that it is research that takes the idea of a "society where robot characters are friends" seriously.

I often use the example of *Doraemon*. I like *Doraemon* quite a bit because I was born in the 1980s. *Doraemon* is very fun, and although I am digressing or going off on a tangent, various artificial beings do appear in the story. *Doraemon* appears to be more like a human being, but other beings that appear in the story are relatively diverse, e.g., a self-driving car with its own personality in one episode and a robot from a completely different planet in another episode.

In this context, I think that what makes us identify agents as being robots is that it is possible to design them. There is a rather famous scene in *Doraemon*, where the protagonists think that if they catch a robot that is their enemy, they can remodel so that it becomes their ally; they say, "it is a good idea," and there follows a very famous line: "We just need to remodel the circuit to the one we are familiar with." I have to say, however, that I find the scene of the two robots agreeing with the humans on that to be rather dystopian.

In short, it is not usually possible to design others, such as humans or pets, but it is possible to design others that are artificially created. Since it is possible to modify them according to the convenience of humans, the point of human–agent interaction is to explore what we can actually do to achieve that and how far we should go with that in the first place.

Historically, it has become necessary to consider these issues since the 1950s, when computers with internal states were created. In my opinion, when computers reached the stage at which they interact with humans, rather than simply doing calculations, it became necessary to do research that considers how to design things that behave with intentions.

It is important for us to step into that area somehow, even though there already is another branch of AI research that considers what is to be done in order for machines to have intentions. While we do not step into this area explicitly, we must consider the impact that it has on humans when machines appear to have intentions. I believe that we need to consider this because it is a relatively practical issue.

Several researchers have recently entered this field, and they can be broadly divided into three types. One is robotics researchers, that is, researchers who have created so-called “humanoids” and “humanoid robots.” They research social robots and have entered this field from the perspective of what we can do with humanoid robots and how far we need to go to make them effective.

The second type of researcher conducts research on conversations. This type of research has been conducted for a long time among AI researchers and has expanded with the introduction of smart speakers, such as Amazon’s Alexa or Google Home. Research that attempts to solve problems through conversations as in the case of smart speakers and such like has been around for a long time. Among these developments, it seems that people who have been thinking about the effects generated not only by conversations but also by emotional expressions and facial expressions have recently entered the field of HAI.

The third type of researcher works in fields such as psychology or cognitive science. Especially those who are in the field of cognitive engineering, or what is known as “applied psychology,” are interested in this kind of study in that it considers ways in which to reproduce the way HAI affects humans with artificial devices.

HAI is about designing something that behaves as though it has intentions. Leaving aside the question of whether it does have intentions, we are attempting to design something that behaves *as though* it has intentions, but we need to think about how to design such a device and what background it will have in the first place.

It is extremely difficult to define what "human-like" means, and I cannot get into this too deeply, but I would like to mention some examples of research in which something similar to what is regarded as otherness in HAI is studied.

For example, there is a definition by the philosopher Daniel Dennett, who published the book *The Intentional Stance* about this.

He says that there are three ways to view the world. Some say that there is actually a fourth way, but let us say that there are three for the moment. In a nutshell, when humans observe the outside world, we have what he calls the "physical stance." That is to say, there is a physical law behind the movements. For example, an object falls down when it is released from someone's grasp, and a round object keeps rolling when someone pushes it.

Children do not exactly know the law of gravity, but they know that an object will fall down if they release it from their hands although they would not know things such as how fast it would fall. They have empirical knowledge. For example, they know that if they spilled water, it would spread out and not return to the container. Dennett says that this way of viewing the world is the physical stance.

However, not all humans view the world in this way. For example, when an alarm clock goes off, we react by thinking that we need to get up. Of course, there are mechanisms such as springs wound inside the clock, and they move like this or that at certain times, etc. However, humans can understand what is going on without interpreting the situation that way. That is because humans can remember a set of functions such that we know a certain function will occur with a certain action. The design stance is the way in which we see the phenomena of the external world, and most artificial products are made with this in mind, for example, it is in knowing that if you press this button, a machine will start working.

Smart artificial devices, such as smartphones like the iPhone, work under the intuitive rules or the physical stance—if we do this, it will expand or shrink—and the design stance—if we press this button, this part will open—which are combined to produce these smart devices. The design stance involves these areas.

However, there is another way of viewing the world. For example, if I was late and someone came to wake me up, saying, “Get up,” unlike in the case of the alarm clock, I would not think that this person was designed to wake me up at this time. Instead, I would interpret the situation by considering the inner aspects of the person, such as thinking that this person woke me up because they were worried that I would be late. Basically, in these cases, we interpret the situation. This happens with interactions between humans or between humans and pets. For example, if a dog comes to me and barks to wake me up, I would not think that this is because it is designed to wake me up at this time in the morning but perhaps because it wants to play. We consider inner aspects in these situations.

Usually, this applies to humans and animals. Although it does not apply to artificial products, the point is that it could. For example, devices such as Sony’s AIBO are designed to make humans imagine exactly this type of situation.

When we talk about these topics, people tend to think that we are exaggerating or that there are not so many anthropomorphic things. However, what becomes relevant here is the study of the media equation, or the so-called “HCI” in Japan. This is research that compiles cases showing that what is affected by these issues is not just the content but the media itself. A famous example is information search using a computer.

Specifically, think about something like a travel plan.

It is just a piece of computer software, nothing special. The research is conducted to test how easy it is to use average software, where the participants use search and command functions and evaluate afterward. We divide the participants into two groups and ask one group to evaluate the results of the search with the computer. (figure 6)

The other group is asked to leave their desks after they have finished with the task and to go to another room to evaluate the results using the same computer in the other room.

Theoretically, the two groups should receive the exact same evaluation; most of the situation is the same except for the location of the computer. However, that is not the case.

The first group receives a better evaluation than the second group. This becomes interesting as we investigate further. Because of the poor evaluation of the second group, we begin to see that their evaluation was more accurate than that of the first group.

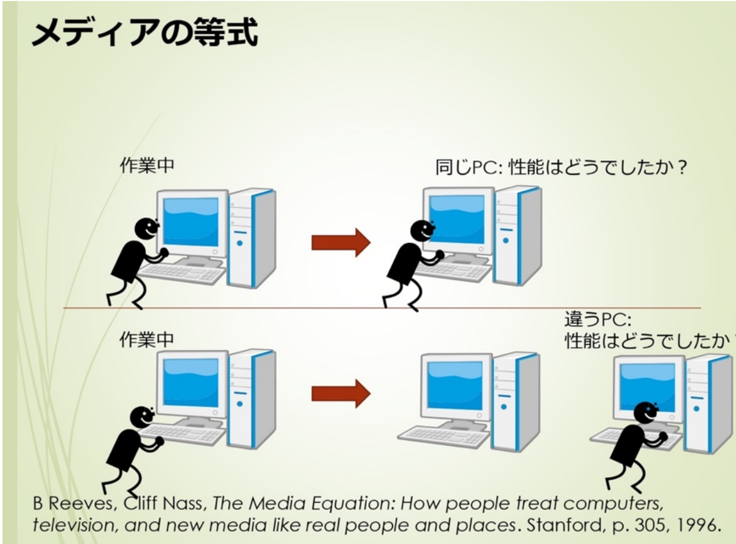

Figure 6

To determine the reason for this difference in accuracy, we conducted the experiment in various ways so as to establish whether the location was the problem. The results indicated that the location was not the problem; the outcome was the same whether the results were evaluated on the same or different computers.

Instead of determining why the evaluation was different, we thought that it would be easier to come to a conclusion if we considered how things would work between humans. For example, after working together, a survey would be completed, where colleagues would ask one another face-to-face "How was it?". The reply would be "It was quite good. Thank you." However, if I asked the colleague who interviewed the other, "What were they like?", the reply would be "I am afraid what they were saying did not quite add up," because it would become a bit easier for the colleague to say so. The same effect is observed on computers as well.

Was the problem that the users were beginners? We believed that we would not see such an effect if the user was someone who was more familiar with the system, for example, a programmer. However, the same effect was observed regardless of the level of computer literacy. In the case of a computer expert, the person would become irritable if asked, for example, whether they had tried to be "nice" to the computer.

The above experiment shows that either humans are socially programmed to feel otherness or to anthropomorphize others much more than originally thought, or such programs are perhaps innate. Researchers call these "social actors" and regard them broadly as artificial objects that can influence society.

This can be related to a theory in biology called the brain hypothesis, which explains the reason for which humans have a brain that is the largest of all species relative to body size.

Another theory in biology explains that humans are herd animals. Other types of herd animals, such as bees and ants, also form a group. However, these herds are usually made of blood relatives. In other words, they are made of individuals that share the same genes.

It is possible that the tendency to form a herd with fellow individuals by genetic rules is an innate trait. For example, while it is possible to pre-program under the assumption that the genes can survive if there is cooperation so that any offspring can survive, there is a possibility of betrayal by those who are not blood relatives but come together for a shared purpose—as is the case with this group. There is also a possibility that somebody will break away from the group of their own will or that the winner will keep the reward all to themselves. It is very difficult to prevent such events by genetic methods because human behavior is unpredictable.

Then, it becomes mutually valuable to learn or to base a model on another.

This is particularly the case with primates, in whose societies, it is necessary to have not only a model of the other but also a recursive model that considers what the other party thinks of oneself. This means operating the other party's simulation in yourself, which becomes a complex task.

Why do monkeys have such big heads? Gathering food is not such a difficult task; one researcher has stated that while living in the jungle is an easy "game," the hard "game" is forming friendships and working in collaboration with others.

In short, humans have a brain that can perform certain tasks, such as reading with intention, that are taught by others and are considered to be something not innate.

A study applied this balance theory in psychology to agents (in other words, to artificial objects).

Although it might sound like a slightly dangerous study, considering that a robot causes humans to quarrel with each other, two friends (participants) and a robot are studied together. The robot is instructed to talk to only one of the participants enthusiastically and completely ignore the other.

The attitude of participants is measured before and after the experiment. In psychology, the balance theory explains the balance in relationships if the multiplication of the negative and positive relationships balances out to a positive; these are then stable.

For example, if every participant has a good relationship with the others, the relationships are all positive and therefore stable. If two have a good relationship, but one has a bad relationship, the situation is unstable.

However, if person A and person B have a good relationship, but person A and person B both have a bad relationship with person C, the situation is still unstable. In short, it is a mechanism that does not create a form of "the enemy of my enemy is my friend." Although this situation is unstable, it can become stable if either the positive or negative relationship changes.

If we reproduce this situation with a robot, we can forcibly produce a positive relationship on one side and a negative relationship on another. However, how will the remaining relationship become stable? We have to convert it to a negative relationship for a positive balance to be attained. Therefore, in this study, as the robot cannot change its attitude, its human counterparts have to and can experience falling outs as a result.

When the participants were asked "Was this robot clever?", they would say "No, it never gives right answers, and it seems to have been merely programmed to say those things," which indicates that there may be jealousy toward the robot.

In other words, the problem is that when participants talk to the robot in a conversation, and although participants question whether the robot is nothing but a computer, humans display a tendency to become jealous when they see the robot and another participant having a friendly conversation. In recent years, research on robot agents has leaned toward the method of investigating communication models between humans in society rather than simply dialogs with problem-solving robots as we have come to believe that relationships with robots may become substitutes for human relationships.

There are some specific examples, such as Dr. Michio Okada’s HAI research at Toyohashi University of Technology leading to his recently published book titled "Weak Robots."[1] (figure 7)

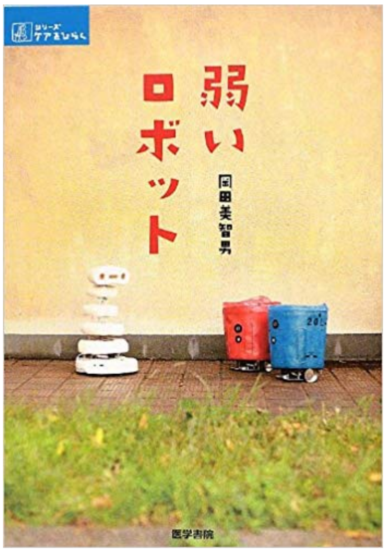

Figure 7

His approach is to question whether, rather than simply creating strong robots, humans can create robots that prompt acts of assistance from humans, for example, by showing that they cannot do anything by themselves.

This is comparable to the Sociable Trash Box (figure 8), which is a trash box that moves but cannot pick up any trash although it tries diligently. It has been observed that when kids see this, they rush together to help put the trash in the box. It is said that the true value of the box lies in prompting human action.

Of course, some would say that if it is simply about picking up trash, why not just use a machine like Roomba? However, that is not the point of this activity—trying to induce human action, as evidenced by the children who actively start picking up trash.

Figure 8

Another example is Pekoppa, created and released by Sega Toys. Dr. Tomio Watanabe is developing this

[1] 岡田美智男『弱いロボット』医学書院、2012 年。（＝OKADA, Michio. 2012. *YOWAI ROBOTTO* [Weak Robots], Igakushoin.）

product and has been conducting research on the concept of "nodding." He has been continuously making virtual agents and robots that nod. For example, in one of his projects, he placed a row of robots at the back of the classroom to study the manner in which they would help a human communicate more easily by simply nodding during interaction. This demonstrates the effect that robots have on society rather than simply demonstrating the intelligence they may possess.

Figure 9

Another example is Paro, which is a seal-type robot created by AIST and was designed to minimize users' sense of discomfort by deliberately setting up a seal as an agent, which nobody had ever seen before. There is also research that examines the differences between humans and agents using Geminoids, which are "human surrogates." This is conducted in the lab of Professor Hiroshi Ishiguro.

The lab of Professor Fumihide Tanaka from the University of Tsukuba conducts studies that use the method of learning by teaching, whereby children teach English to a robot. This aids the children's own learning such that they can learn the language while teaching it. One believes that this demonstrates the recent trend of HAI research—to make changes while addressing a wide range of topics.

We have talked about cyborgs, and I took some time to think about the situation "after cyborgs," as well. If the function of a type of cyborg is to expand human evolution, perhaps the move toward artificially expanding human society itself is the direction in which the technology of HAI advances. Humans have to consider the possibilities that this can offer.

One could create things similar to Geminoids, but instead, I have been thinking about creating something complementary to human society, by experimenting in various ways and trying to brainstorm ways in which to utilize the actions that are unique to each agent.

In this area, the concept of "synthesizing otherness" is highlighted. (Figure 9) At first glance, it may look like an ordinary robot that is moving and engaging in discussions with others. However, this is not an AI but a remote-controlled robot that is maneuvered by a person. There are many studies that focus on remote presence as an approach that rather than creating a feeling of distance—as in the case of a teleconference—allows for more realistic robot use, with the incorporation of facial expressions.

However, there are also fake actions involved. These are not all actions of the user; merged with the user's actions are those of the person with whom they are talking. The key aim of this research is to make the other party feel more comfortable and to create an atmosphere of effective and easy conversation (Figure 10). The

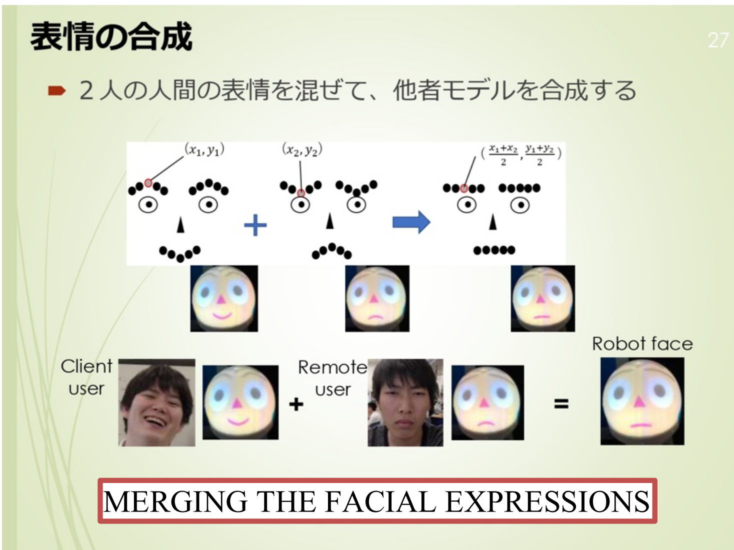

Figure 10

mechanism traces both faces and merges the facial expressions that are displayed.

For example, if the party in a remote location is laughing, and the local party does not feel like laughing, the robot synthesizes the emotion by weakening the laugh.

The robot reflects the behavior of the client, and the facial expressions are merged to keep a balance that is based on the client's emotions. The condition on the left is a pattern that completely projects the behaviors of the remote party. Two conditions are prepared: mirroring the behavior of a person in blue and the behavior of another user. How the impressions change under the two conditions despite having exactly the same conversation is assessed. (Figure 11)

Figure 11

I will not go into the details of the conditions, and the results are nuanced. Simply put, we investigate how people agree or disagree with each other by having two participants discuss a problem called the "desert survival problem" through a robot. Afterwards, participant A is queried on the satisfaction level of the interaction with the other person, participant B. The results show that participant B, who is in a remote location, becomes easier to understand and therefore more positively evaluated if the robot mirrors the movements of participant A.

This is the result when participant A does not realize that the robot is actually copying their actions; if they do, the evaluation becomes markedly lower.

The research was described in a critical manner because the task given to these participants was to have a simple, general discussion about, for example, social acceptance, which made the robot seemingly a little more effective as a result.

Another study, which originated from a desire to surrogate the emotional labor of humans, shows the eyes of the other instead of human eyes (Figure 12). This concept is defined by sociologist Arlie Hochshield circa the 1980s. She argues that there are three types of human labor: physical labor, intellectual labor, and emotional labor. Emotional labor is not a process of calculation or search in particular; it is a process whereby a person controls their emotions in accordance with the other party.

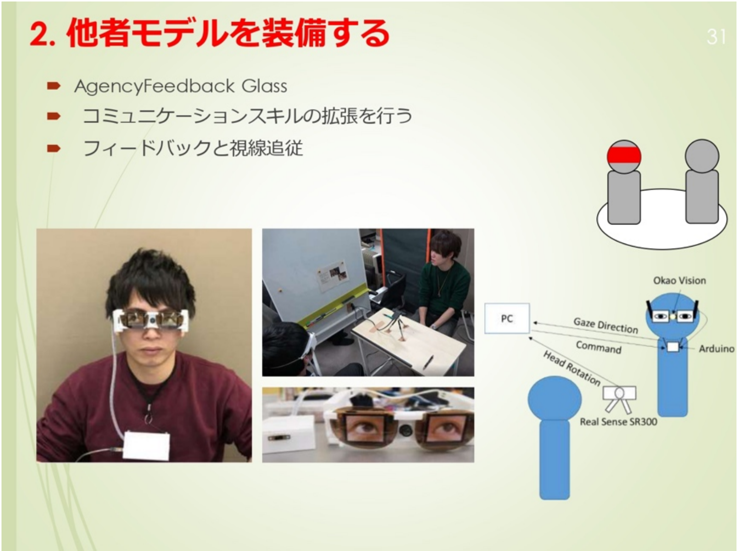

Figure 12

As a technology to assist with emotional labor, an artificial gaze that can express consideration for the user has been created. In the present situation, it may be effective as an interface for blind people, fulfilling a more practical role. While there are many technologies to help the blind, they are unable to communicate with eye contact, which is something that they feel frustrated about, according to preliminary data collected from a first survey run. A model was used to discover how to solve that. The main issue is regarding whether the robot can meet the other party's gaze by following it. As an example, if one wanted to follow another person's gaze, to look at what they are looking at while having a conversation, the technology

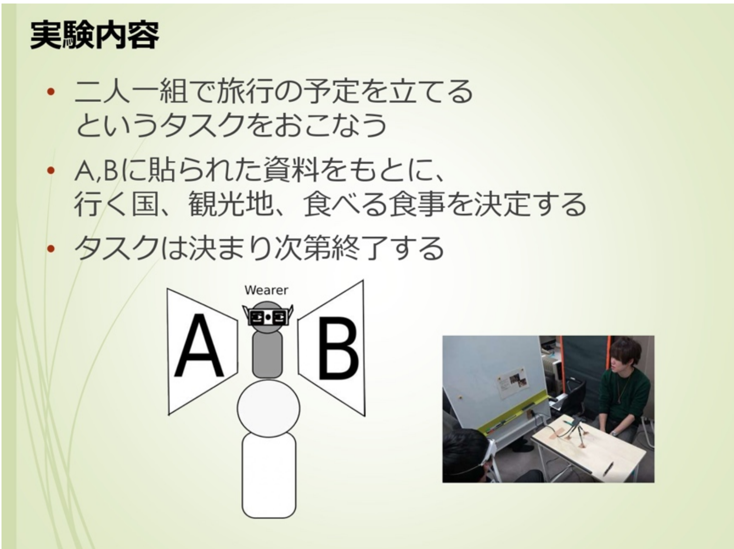

Figure 13

would be highly effective there as it would enhance the feeling of the experience.

An experiment was conducted with nondisabled people; they were placed in a situation where they could not see each other, and they were asked to have a conversation about planning a trip while wearing a device (Figure 13).

The device detected whether they would turn their faces after receiving the information on what the other party was looking at and whether they would follow each other's gaze.

From the viewpoint of performing the task with ease, it was concluded that it would also be better for the other party if they had feedback from the person wearing the device.

The device looks at the other person's gaze and moves its gaze in accordance with it, and as the information is conveyed, it results in the movement of the wearer's neck. Although one has to perform the action by oneself, it is observed that this act of moving the neck makes the task easier to carry out. Also, an overlap of voices occurs when two people utter words at the same time. In short, when a conversation is developing smoothly, if they say something like "Oh!" simultaneously, it undermines the smooth development of the conversation, but a definite decrease of the overlap rate has been observed.

This way, although it is a very modest element, the development of smoother communication has been noticed, which makes this study pertinent. Indeed, the most important function of the eyes is to see, but a certain level of artificial assistance can be provided so as to aid communications with others.

Recently, one of the most prominent areas of HAI research has been the influence of the others’ presence, called "social facilitation." (Figure 14) An example of this effect is when one is happy if someone is with them while they are working. That is simply the name that has been given to this sense of earnestness. A robot's feeling the presence of someone, by itself, can motivate people to perform better. This research examined how significant that effect is.

This experiment specifically focused on eye movements, which essentially means that the participant was tested on whether they would see the same thing as the other party. This study examines the effect of a joint gaze called "joint attention."

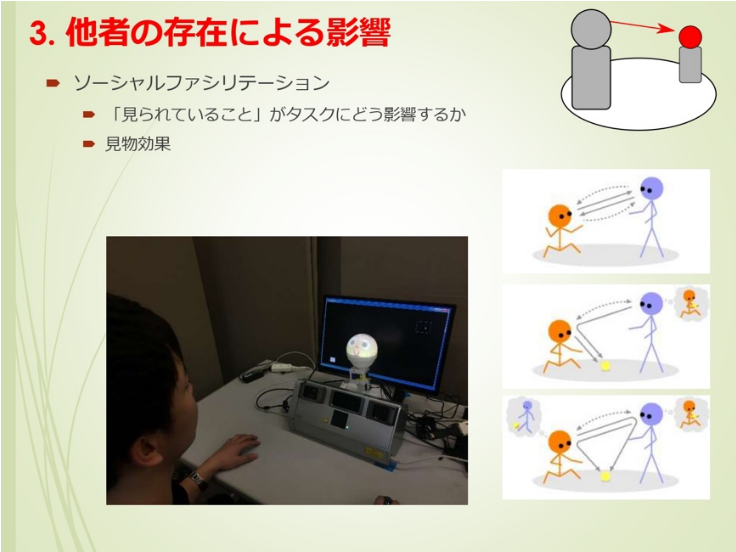

Figure 14

In fact, among extant research, researcher Dr. Masahiro Shiomi of ATR has been studying the effect of being touched by a robot while performing a task. This is a very cheery and pretty experiment, so to speak; it involves the robot gently rubbing the back of one’s hand while they work. Although the effect is very subtle, the study shows that the act of touching the hand of the participant while they work has a positive impact on their task performance.

This cannot be directly replaced by a gaze. We have experimented with gazes as they can be applied in many situations since, unlike touching, they do not have restrictions. However, such use of gazes did not result in improved task performance. However, it has been found that gazes can considerably improve people's impression of someone or something. For example, we have observed a tendency toward improved motivation and concentration if a robot moves its gaze not only at random or toward the other party but in accordance with the gaze of the other party.

Another aspect is that examining how often the participant was looking at the robot during the experiment, it seems that it can be indirectly confirmed that the participant was more concentrated on the task, as they turned to the target more often, when the robot followed their gaze.

Studies arguing that a robot's presence alone can be effective have also been attracting attention.

Last, a study that was conducted at an elementary school examined what would happen if we directly designed the agent ourselves.

Specifically, this is a robot that introduces books. It appearance has changed now. This robot displays the face that the children designed for it here, displays introductions of books prepared by the children themselves, and can talk to the children. It is difficult to see, but it changes facial expressions, explains the content of the book, and enunciates whether it was interesting to read while making gestures. (Figure 15)

Figure 15

The aim of this study concerns the area of so-called "user-generated content," which has become popular, notably on Youtube. This means that while content was formerly made by professionals and watched by the general public, the emergence of this new idea means that content is made and evaluated by the users themselves and that they produce new content in response to the evaluation.

Simply stated, the study applied that process to the agent. In the past, contents of the agents, including products like AIBO, for example, were essentially made through programs designed by skilled people and enjoyed by everyone. However, there have been problems with this method, such as the fact that people easily get bored with the robot or that people do not use the robot over the long term.

The most interesting part of the new method is the creation of the agent by the users. This study opened up that part to the children and returned the data, such as the titles that the agent displayed or how many books the agent handled or read out for the users. The goal is to create a cycle in which children who would not usually read books begin to read and develop their ability to express themselves further.

Our research makes use of agents such as this and inserts agents into society but with an approach that is somewhat different from what people are used to.

Specifically, we design programs such as this even as we implement them. (Figure 16) The point is that although these kinds of agents exist in abundance, because the agent is made by children themselves, it is easier for them to understand it. Thus, their opinions are considered, and the program is designed to be tailored to children. For example, it is easy to use and has movements and emojis, adjustable timings, and even colors.

Figure 16

After creating the contents, children can also check regarding the number of other children who have listened to their explanations of the books. The program is designed to consider the situation in which they want to continue to change the explanations so as to improve the contents by altering certain parts because, for example, few people have read the explanation of a book despite its being very interesting to the child.

Since this is an ongoing study, not much can be concluded, but there was a correlation between this type of agent and its impact on the children. A very basic result is that a correlation exists between the number of times books are introduced to children and the total instances that they picked up the books.

However, the content remains limited. The workshop is being conducted by explaining the robot that we are making right now with an improved body.

In addition, the team is engaged in many other things. It is being considered whether it could be possible to deliberately detach our bodies from ourselves by turning them into agents. As a very simple example, one may be able to stay on a diet longer because their right hand is still trying although they have given up. The feasibility of such research is being assessed. The task itself remains the same, but it prompts another physical behavior that is unrelated to the task, and this is being investigated.

Moreover, there is a study involving an agent that eats food together with the user. (Figure 17)

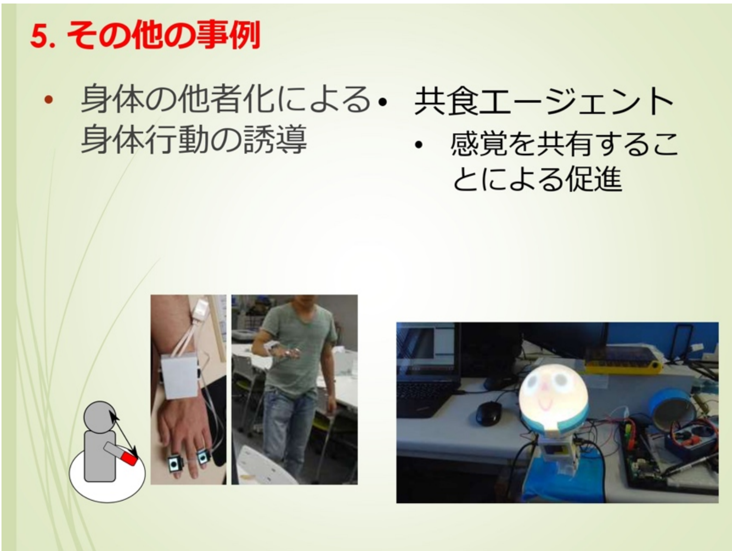

Figure 17

Indeed, there is a field of study called co-eating. In short, it premised on the study of how to educate children about eating. One of the projects is to see whether results can be obtained by producing an impression that the agent itself is eating. Right now, this is being carried out by adding eating motions to the agent to see how it can help children eat.

Another project is the aforementioned mafia game. (Figure 18) I am conducting this research with various researchers. To describe this simply, it is a game to discover spies, and we study ways to make AI spot a traitor during communication. There are various approaches to this, but one of the actions is to model this game.

Particularly because I am specifically interested in the agents' communications with humans in the field of HAI, discussions are being held on how far the communications can be modeled to create simulations or to make the formula feasible. Another action is to actually create a robot agent that makes motions and to examine the effects of its facial expressions or motions as someone plays the mafia game with the robot while they have a conversation.

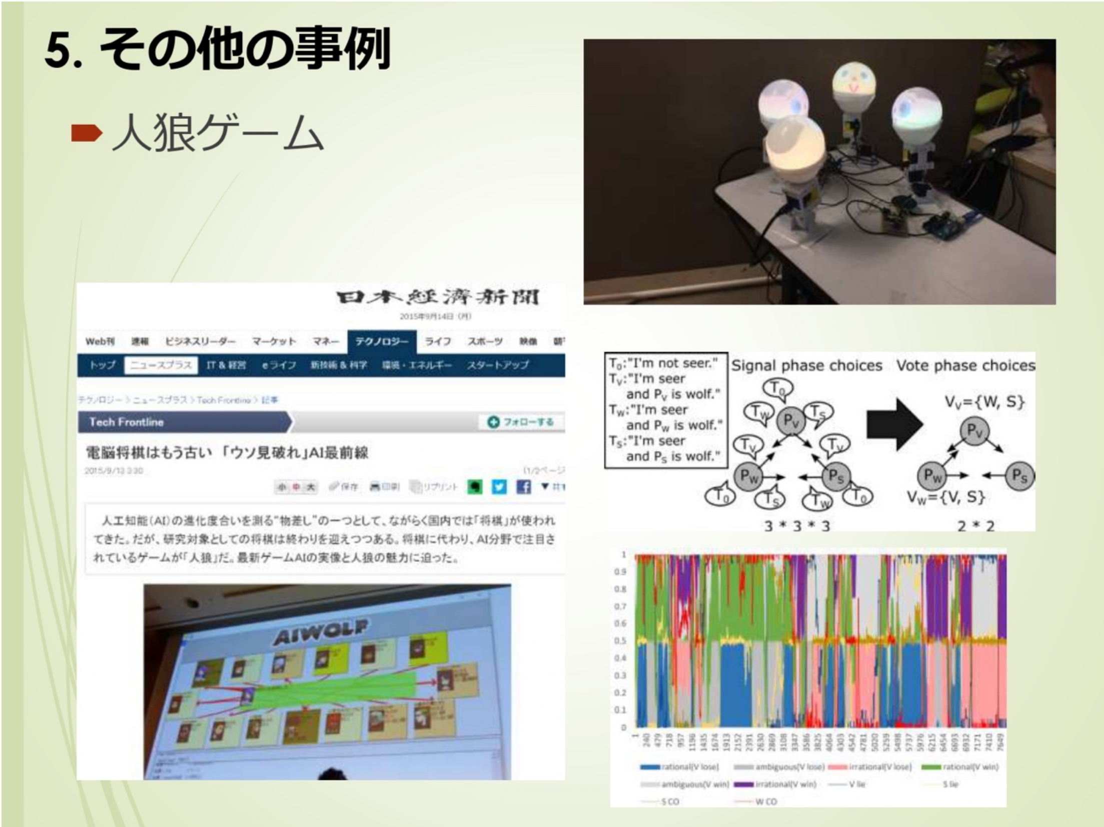

Figure 18

Specifically, we discuss how AI agents learn from one another the human behavior of reading the meaning behind words even when they are in an environment where communication is very limited.

In addition, for example, in the card game HANABI (Figure 19), players are not allowed to see their own cards. This game is played by cooperating with other players and attempting to guess what others are thinking. I aim to examine how to concretely reproduce this act of guessing by using an algorithm. I believe that the results of the game can be improved by creating simulations of the players' thoughts on the basis of their gaze.

Finally, I would like to discuss the future of HAI research.

I have also got this from Doraemon; it is rather ironic because I like Doraemon.

It is necessary to stop and reflect on whether it is worthwhile to make friends with a machine or rather to make a machine-like agent.

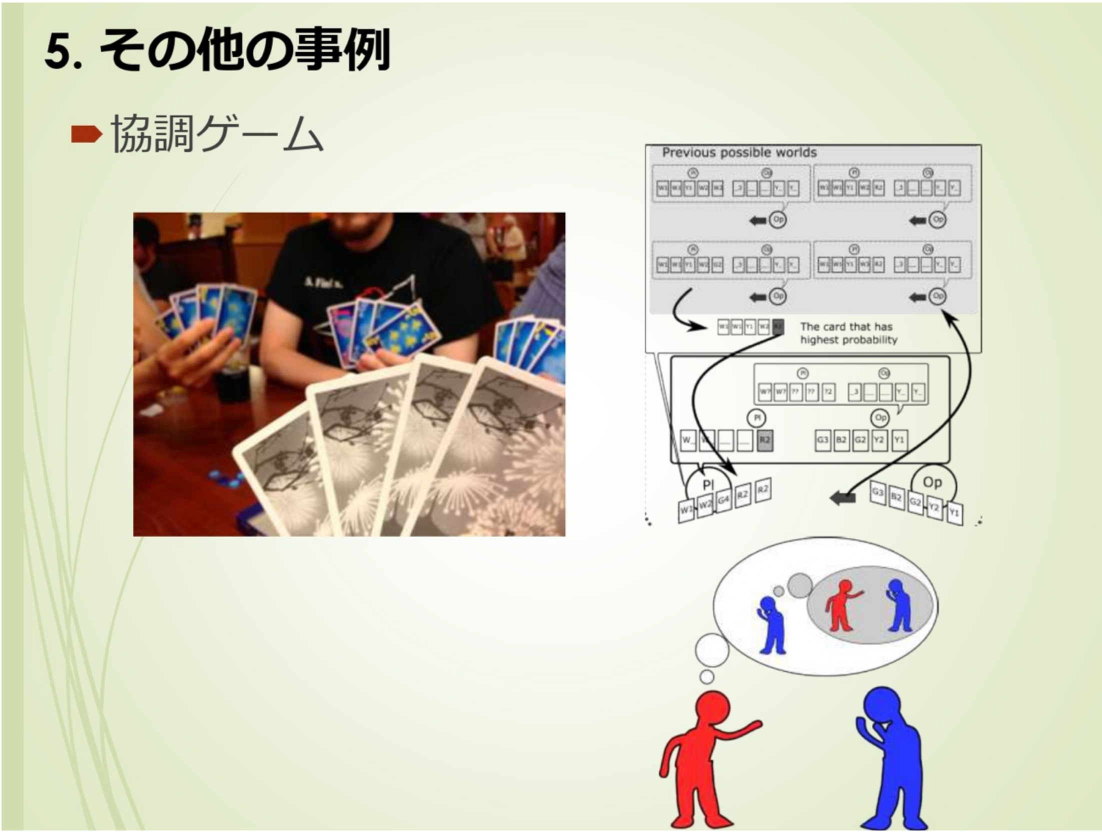

Figure 19

When arguing from the perspective of studies related to the history of life, this is often mentioned: life did not choose a tactic that allows only genes to determine everything as creatures like mammals learned how to behave more constructively in accordance with their environments. From such a situation, creatures like humans were born. They leave behind an environment externally by using records, languages, and cultures to increase the knowledge to be passed on.

The world today is the result of these creatures' decision to build computers that can hold memory data, to build robots that they can recognize as "others," and to create interfaces for these robots.

These can influence people at least to some extent. I believe that this method will develop more and more because there are some useful situations in which, for example, social facilitation can be worked on with the agent on the basis of studies such as the aforementioned one about making people jealous of the relationship between the user and the agent.

What would happen when the method develops? As stated many times, I have a passion for studying Geminoid. Nonetheless, I do not think android-type robots are the ultimate form. For example, it is not necessarily useful to have robots standing in various places in the company building because robots are not demanded to do the same things that humans do, given that what we have by nature is different from the things that influence others.

This is a very familiar example: sparkling eyes in an animation film or a comic book are not at all like real human eyes, but I believe that these are selected because humans choose the most suitable ones according to their perception of the work.

If that is the case, I believe, for example, that screen agents and robots will come in a huge variety of designs, some of which may have the ability to perform tasks better than humans would, and this requires further discussion.

Indeed, many more possible ways exist to place robots, particularly regarding complementarity, which involves matching tasks. The reason for which optimization will happen in many ways is that considerable competition is present.

This is a video of a model that Apple made in 1987 after Steve Jobs was fired from the company. This model has a secretary-type computer in the upper-left corner, and the user asks it for advice. This dream was realized by the voice response of the iPhone when Siri was released in 2011, but actually, during this time, it had gone through considerable changes in design to improve it. For example, Siri's expressions are the result of the reassessment of whether they are really sensible for the tasks.

Conversely, as an example with Siri, it initially spoke only with a female voice, but as it was judged to be culturally inappropriate, a male voice was added in order to add more variety to the expressions.

With regard to this, an agenda is beginning to emerge, which is to make AI agents for them to not simply provide good performance in terms of engineering but also to identify whether society accepts them.

For example, it was decided that a woman's voice would be better because both men and women preferred a woman's voice, according to a survey conducted by the people in the field of HAI. That is probably why Apple adopted the view without question, but it is perhaps slightly out of the boundary of social acceptability. It is considered to be desirable if the users can select either a female or a male voice, considering social equality between men and women. Knowledge of those areas will be increasingly useful for AI agents to be accepted into society.

Some related studies are beginning to enter the field of HAI, such as the research on HAI, AI and multi-agents, social networks, psychology, and medicine. I wrote this particularly from the perspectives of the humanities, applied philosophy, representation studies, and ethics as they will also be relevant in the future.

Three years ago, our society had a picture like this on the front cover, and it caused a fuss (Figure 20: right above image). This was discussed by writing papers, but with

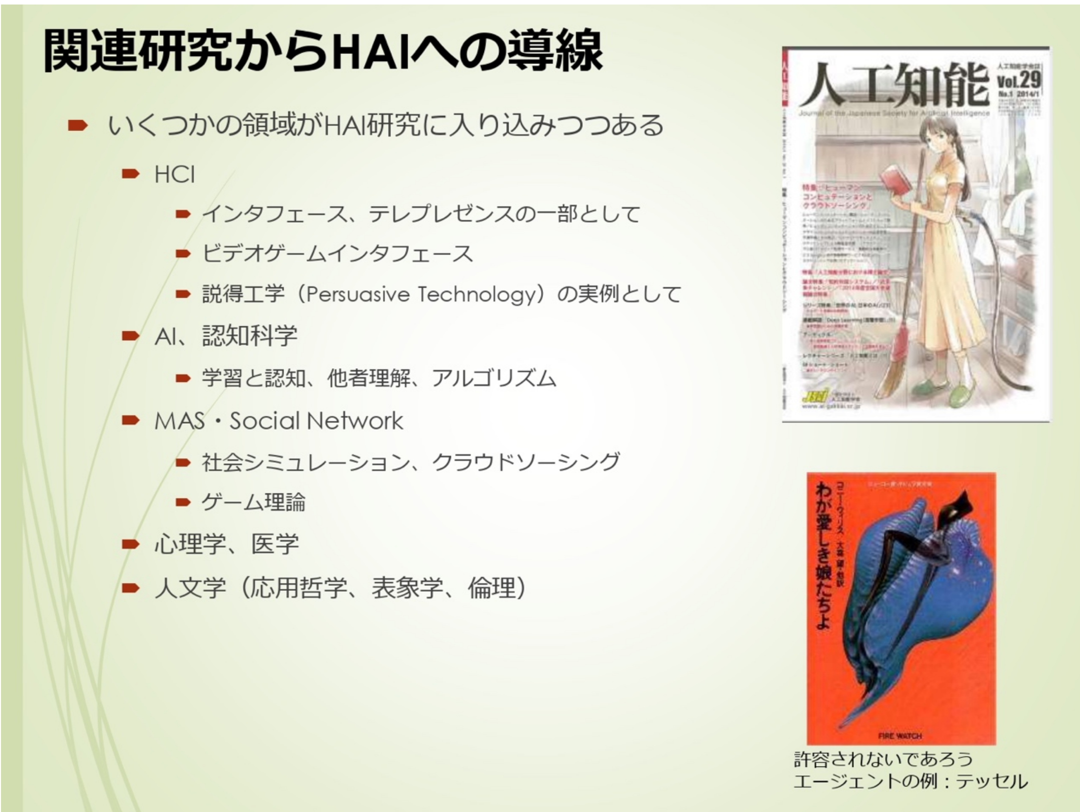

Figure 20

regard to the matter, we were especially told that we were studying anthropomorphism and asked to write explanations. The idea that AI agents should not be simply designed can be widely shared as there are several points to be considered. We argued that it was not a serious problem because it was merely a picture on the front cover, but it would be problematic to actually design something like this.

In particular, this is present in Donna Haraway's aptly titled *A Cyborg Manifesto*. She mentions agents that remind people of child abuse, and although there they are creatures from outer space, such agents could really be designed. Research has been conducted to create a robot that incites bullying, and there is a study to identify a process on how an existing robot is bullied. However, we will have to consider whether those robots are acceptable as a separate issue from the impact of such studies.

Particularly in Europe, some argue that robots have different impacts, depending on their skin color. Sometimes, robots become unacceptable if they have the "wrong" skin color. However, such effects should not be applied regardless of whether they will contribute to better performance, and I strongly believe that we need to have discussions with the researchers in the field of the humanities on that point.

In the future, research on agents will probably have to involve not simply the design of agents but also the design of interactions, and therefore, the job will be something very similar to designing society.

Such a design has been described in old science fiction stories, and especially with ASIMO, but I feel that those areas will become more realistic.