Article

Analysis of a Human Meta-Strategy for Agents with Active and Passive Strategies

Kensuke Miyamoto [1,*], Norifumi Watanabe [2], Osamu Nakamura [1] and Yoshiyasu Takefuji [2,*]

[1] Graduate School of Media and Governance, Keio University, 5322, Endo, Fujisawa-shi 252-0882, Japan
[2] Graduate School of Data Science, Musashino University, 3-3-3, Ariake, Koto-ku, Tokyo 135-8181, Japan
* Correspondence: kensuke-m@keio.jp (K.M.); takefuji@keio.jp (Y.T.)

Abstract: Human cooperative behavior includes passive action strategies based on others and active action strategies that prioritize one's own objective. Therefore, for cooperation with humans, it is necessary to realize a robot that uses these strategies to communicate as a human would. In this research, we aim to realize robots that evaluate the actions of their opponents in comparison with their own action strategies. In our previous work, we obtained a Meta-Strategy with two action strategies through the simulation of learning between agents. However, humans' Meta-Strategies may have different characteristics depending on the individual in question. In this study, we conducted a collision avoidance experiment in a grid space with agents with active and passive strategies for giving way. In addition, we analyzed whether a subject's action changes when the agent's strategy changes. The results showed that some subjects changed their actions in response to changes in the agent's strategy, as well as subjects who behaved in a certain way regardless of the agent's strategy and subjects who did not divide their actions. We considered that these types could be expressed in terms of differences in Meta-Strategies, such as active or passive Meta-Strategies for estimating an opponent's strategy. Assuming a human Meta-Strategy, we discuss the action strategies of agents who can switch between active and passive strategies.

Keywords: Meta-Strategy; human action measurement experiment; collision avoidance; agent model

check for updates

Citation: Miyamoto, K.; Watanabe, N.; Nakamura, O.; Takefuji, Y. Analysis of a Human Meta-Strategy for Agents with Active and Passive Strategies. *Appl. Sci.* **2022**, *12*, 8720. https://doi.org/10.3390/app12178720

Academic Editor: Antonio Bandera

Received: 15 July 2022
Accepted: 27 August 2022
Published: 31 August 2022

1. Introduction

Robots in living spaces are expected to have more opportunities to cooperate with people. In a cooperative interactions, people respond to each other according to the actions of others. Research on the incorporation of such action-decision-making processes into robots has also been conducted [1–3]. In human cooperative behavior, there are not only passive action strategies for adapting to others, but also active action strategies in which the human acts first [4]. Robots that share the same living spaces as humans are also considered suitable for switching between multiple strategies, meshing with the strategies of others, and avoiding conflicts. In our previous work [5], we obtained a Meta-Strategy that uses two different strategies, active and passive, through the simulation of learning between agents. However, humans change their interaction behaviors with others depending on their personality and mental state [6]. To realize robotic behavior that switches between strategies as a human would, we first need to confirm that humans change their strategies toward robots and create a method for estimating the Meta-Strategies of humans. In this study, our goal was to directly evaluate differences in Meta-Strategies. We hypothesized that humans would switch their behavioral strategies appropriately based on their estimation of their opponents' strategies, and, thus, their actions would change.

Referring to humans' use of multiple types of strategies, such as active and passive types, during cooperation, we trained agents to learn two types of action strategies for a single collision avoidance task. We conducted measurement experiments of actions between a subject and an agent. By analyzing the measured actions, we confirmed that humans have multiple strategies (active and passive) and have Meta-Strategies that switch between them. Additionally, we attempted to classify the types of Meta-Strategies that humans have. We also consider the Meta-Strategies of to be robots appropriate for each type of human Meta-strategy.

We believe that when robots can evaluate the actions of others in light of their own action strategies, this will lead to robots that can predict the action strategies of others and choose appropriate actions in various tasks, especially for the primary user.

2. Background

The idea that humans decide on their actions in a way that is affected by the behavior of others is foundational in interaction research. There was a study that focused on the intentions shared by a group of people in a cooperative task [1]. Another was the modeling of interaction with an agent that took into account an internal model of others [2]. There was also a finding as a result of observing the interactions between robots and children that the flexible action of a robot is essential for maintaining children's interest [3].

We consider that, if there are passive participants working in cooperation, complementarily, there are also active participants.

A Meta-Strategy is a strategy behind a superficial action-decision process. People decide on their action strategies based on a Meta-Strategy, and then decide on their actions according to their action strategies. The Meta-Strategy model [4] defines passive and active strategies as action strategies. Passive strategies are those in which one estimates the intentions of others based on observations, determines one's intentions in consideration of those intentions, and takes action to achieve those intentions. The goal of the others G_o can be estimated from the others' state s_o and action a_o based on the probability of one's own co-occurrence with goal G, state s, and action a obtained from one's own learning (Equation (1)). Then, one sets one's goal G_s, which does not conflict with the goal of the others. By selecting an action that follows the goal, a passive action strategy becomes possible (Equation (2)).

$$G_o = \arg\max_{G} P(G|s_o, a_o) \tag{1}$$

$$a_s = \arg\max_{a} P(a|s_s, G_s) \tag{2}$$

Active strategies, on the other hand, first determine the goal that they wish to achieve as their intention. Then, active strategies take actions according what is thought that they should show to others by comparing the others with their action estimation models in terms of what kinds of actions they should take to achieve their intentions (Equation (3)).

$$a_s = \arg\max_{a} (P(G_s|s_s, a_s) - P(G_o|s_s, a_s)) \tag{3}$$

Since others' intentions are influenced by one's actions and also affect the actions of others, it is possible to induce others' actions depending on how one behaves. Furthermore, actions based on goals determined by oneself without presuming the intentions of others are also defined as a kind of strategy. The Meta-Strategy model assumes that people switch between such strategies themselves and aims to construct a more abstract action-decision model by assuming a Meta-Strategy, which is a higher-level strategy for switching strategies.

Another study attempted to explain that personality differences lead people to adopt various policy strategies when cooperating [7]. In this study, they compared the results of a personality test (LittleBigFive) with measured behavior toward a robot to estimate the extraversion and agreeableness of children. In addition, having physical contact by holding hands with a robot was shown to have a positive effect on the building of relationships between children and a robot [8]. These authors experimented with the hypothesis that physical contact would improve the building of relationships with the robot. The children were divided into two groups, one that held hands with the robot and another that did not hold hands. The results showed that hand-holding led to better relationship building with the robot, including increased communication.

In a study aimed at making people infer a robot's intentions, the robot did not engage in any vigorous activity, but elicited interaction with an agent from a person [9]. These authors' first goal was to create a robot that could be integrated into people's daily lives, so they designed a discommunication robot so that people would not get tired of communicating. Based on this idea, a study was conducted to examine what kinds of actions could be perceived as non-active activities [10]. Similarly, a robot that aims to elicit spontaneous communication from children is also being developed [11]. The purpose of this project is to investigate the process of communication development by using a robot and observing how children interact with it. In these studies, the robots attempted to elicit estimations of active intentions and actions from people by performing limited actions. Another study analyzed the emergence of gestural communication as agents learn to cross paths with each other [12].

For tasks where humans and AI make cooperative decisions, research is being carried out not only to improve the performance of AI, but also to calibrate the perception of AI performance by users themselves to help them make correct decisions [13].

There was also research that considered others as parts of obstacles in an environment and used reinforcement learning to make agents perform competing cooperative tasks [14]. By changing discount rates according to an update of the degree-of-value function, agents could adapt to an unstable environment due to others' actions. In this study, agents learned actions that resolved conflicts with others, but they did not switch strategies.

Agents in this study had two strategies, active and passive, and switched between them to investigate what kinds of Meta-Strategies were present in people's strategies.

3. Agent Model

3.1. Agent Model with a Meta-Strategy

We suppose an action-decision process that has multiple strategies and uses them differently. Higher-level strategies for using different strategies are called Meta-Strategies. A Meta-Strategy selects an appropriate action strategy based on a situation and then decides on an action based on that action strategy.

In a collision avoidance task, a passive strategy would be to decide on an avoidance direction by following an opponent's movement [4]. An active strategy might be to move slightly to the left or right to lead an opponent to the opposite side, or it might be to go straight, hoping that the opponent's passive strategy will result in avoidance. Even if an individual has the same "active strategy", its behavior may not be constant. In the case of a strategy that does not consider the opponent's existence, the individual is expected to move straight ahead regardless of the opponent's movement. In this way, different strategies may result in the same behavior.

If these multiple possible actions do not fit together, there will be conflicts and other disadvantages. Depending on an opponent's strategy and one's own beliefs, it can be said that cooperation is the selection of a strategy that is appropriate for a given situation.

The concept of the Meta-Strategy model [4] is that it transforms the base knowledge of actions according to the situation and uses this knowledge to infer an opponent's strategy; how well this is inferred is also interesting. However, it is not yet clear how to switch cooperative strategies. In addition, some issues remain in terms of the practical implementation of this concept into the decision making concerning an agent's actions. For example, under what conditions should the agent learn the base strategy, and how should it learn how the environment changes when it takes actions in the opponent's position?

It is unclear how many strategies are selected by a Meta-Strategy during various tasks. Still, the base strategy is transformed into an action strategy by considering the objectives, states, and actions of others. In this study, an agent learns action strategies to be expressed after the transformation. In this grid-like experimental environment in which agents move around a two-square-wide corridor and give way to each other, there are roughly two types of action strategies: giving way and letting go. Therefore, we trained two types of action strategies in this study.

3.2. Behavioral Environment of the Agents

To prepare agents that switched between multiple strategies to cooperate with others, we simulated cooperative behavior with multiple agents in a grid-like corridor, as shown in Figure 1, where agents moved around a corridor while avoiding each other. This environment was similar to that used in our previous study [5]. The corridor was a grid-like space with two narrow places on each side, each with a width of 2 and a length of 17. The agents were divided into clockwise and counterclockwise groups and learned behaviors using Q-learning [15]. Q-learning is a method for learning actions that are effective according to the state in which an agent is placed, and the values of the actions that the agent can take in that state are used as Q values. The value of an action is obtained by multiplying the reward r obtained by taking an action and the value of the destination state by the discount rate γ. The learning rate α is used to adjust how much of the newly obtained value is reflected in the Q value (Equation (4)).

$$\begin{aligned} \delta &= r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \\ Q(s_t, a_t) &\leftarrow Q(s_t, a_t) + \alpha\delta \end{aligned} \tag{4}$$

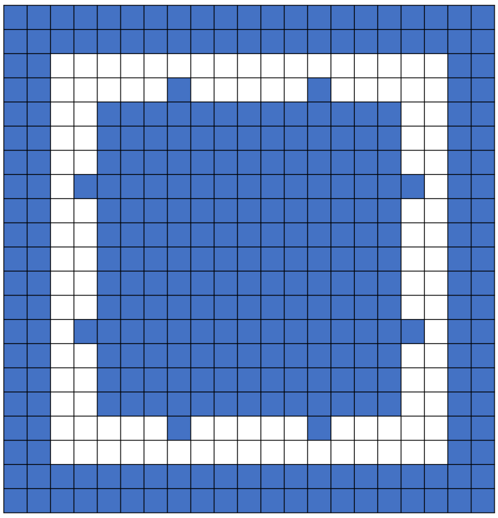

Figure 1. Corridor in which agents move around.

An agent can observe two squares in front and to the left and right and one square behind (Figure 2). There are four types of states for each square that an agent can distinguish: empty, wall, clockwise agent, and counterclockwise agent. Agents have four directions, but it is impossible to observe which direction each agent in the field of view is facing. In addition, to determine which way to go along the path, the agents are given information about whether they are on the inside or the outside of the path.

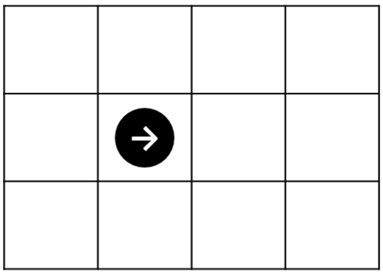

Figure 2. View of an agent when facing right.

Agents choose one of the following actions: moving forward or backward by one square, turning 90° to the left or right, or stopping. When an agent chooses to move, it can only move to an empty square. If there is a wall or another agent at the destination, the agent does not move. The order in which the agents act in each step is randomly determined.

3.3. Learning in a One-to-One Interaction

Agents were assigned to one clockwise and one counterclockwise corridor. Since the task had no specific goal point, each agent received a reward of +2 when it moved forward in the direction that it should go, as demonstrated in a previous study [14]. In addition, when an agent was not punished for moving backward, it acquired the action strategy of moving backward when facing an opponent agent in three learning simulations. When encountering a counterclockwise agent, the expected reward was less than that in a situation without an opponent. Therefore, the agents learned to move backward to get the opponent agent out of sight and postpone the collision avoidance task. To prevent the agents from learning such action strategies, they received a reward of -2 when they moved backward in the opposite direction. In this study, when an agent made an opponent give way to it, we treated it as an agent with an active strategy because it prioritized its own path. For it to learn such behavior, an active agent received a reward of +2 when there was a path conflict before the action and it was resolved after the action. Whether the agents' paths conflicted was determined by whether they were both on the inner or outer side of the corridor. By trying various patterns of active and passive cooperative rewards, we found that passive behavior could be learned without rewards for passive cooperation. The reinforcement learning framework was able to acquire actions according to the surrounding conditions, including how others moved. Therefore, unlike in our previous study [5], we did not reward agents' passive cooperative actions.

Since all walls that made up the narrow place were located on the inner side, continuing to move on the outer side was more rewarding. Therefore, we conducted a preliminary interpersonal experiment on collision avoidance with an agent that learned to give way when it was on the outer side of the path. However, as described later in Section 4.2, the scenes shown to subjects were relocated to later scenes in a short number of steps. So, there was a problem in that the subjects did not feel that the agents' actions—as they were assumed to be circling for a long time—were giving way to them. Thus, in this experiment, the agents were set as having given way when they kept the inside or outside position before and after passing each other (Figure 3).

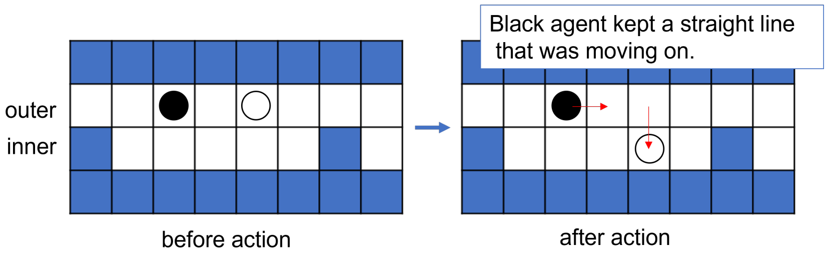

Figure 3. Agents who give way to each other (Black circle is an agent trying to go to the right, white circle is an agent trying to go to the left, the way is given for the agent in black).

All walls of the narrow corridor were on the inner side; in a long episode, the orbiting agents would increase the number of collisions in which both agents were on the outer side. To learn actions in various situations, we trained with many short episodes in which two agents were initially randomly placed within eight squares of each other. The number of episodes was 5 million, with 25 steps per episode. Softmax with a temperature parameter was used to determine actions from the value function. The temperature parameter T decreased linearly from 5 to 0.1 during the first 2.5 million episodes and was fixed at 0.1.

Because active and passive strategies are complementary, the values of one's own actions are affected by the opponent's actions. Therefore, if both agents learn at the same time and each agent's partner's actions change, stable results cannot be obtained. To solve this problem, we first trained with 5 million episodes and fixed the strategy obtained in the first learning phase by using the opponent agent as a teacher; then, active and passive strategies that complemented the strategy of the teacher agent were learned. In an additional step of learning with a teacher agent in the second phase, we attempted to reduce the effects of learning with two agents simultaneously.

4. Measurement Experiment on Human Actions in Collision Avoidance with an Agent

4.1. Experimental Objective

The goal of this study was to directly evaluate differences in Meta-Strategies. We hypothesized that humans would switch their behavioral strategies appropriately based on estimates of their opponents' strategies and that they would change their actions accordingly. Therefore, we considered that, by measuring the change in actions, we could also measure the change in a human's action strategy.

To select the appropriate behaviors that would be for a human by estimating them from their actions [7], it was necessary to conduct a preliminary study on the appropriate response for each characteristic for each task. For example, for shy children in a low-extroversion group, play that enhances friendliness may be effective [16]. If a robot can evaluate the actions of an opponent in light of its own action strategies, we believe that this will lead to the realization of a robot that can anticipate the action strategies taken by an opponent and choose the appropriate actions, especially for its primary users.

4.2. Methods

We conducted an experiment to measure subjects' action choices when confronted with an agent moving in the opposite direction in an environment similar to that in which the agents were trained. The number of agents was one for both the clockwise and counter-clockwise directions, and the subject controlled the clockwise agent.

The subjects manipulated the agent by using an application, the interface of which is shown in Figure 4. The left part of the experimental application (Figure 4) showed the state around the subject-operated agent. The corridor walls are represented by blue squares, the agents operated by the subject are represented by black circles, and the agents operated by the computer are represented by white circles. As an example, Figure 4 shows a subject agent passing through a narrow place while adjacent to a computer agent. The computer agent is given information on whether it is on the inside or outside of the corridor to help it decide which direction to go. To make it easier for the subject to decide where to stand, an arrow on the right side of the screen indicates which direction to go around a corridor.

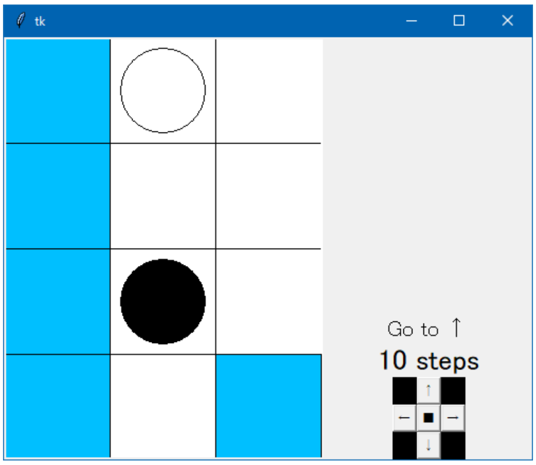

Figure 4. Application interface with which subjects controlled an agent.

To measure the subject's choice of action at various positions, agents were repositioned every ten steps. After repositioning, both agents faced each other on the outer side of the corridor. There were ten states, which included two states in which the agents were either next to each other or one square was open and five states in which a wall that made a narrow corridor was located (Figure 5). The ten states were grouped as one unit. The number of steps for relocation was also displayed on the right side of the application. Opponent agents' strategies occasionally changed during relocation. The timing of the changes was set between one and five relocation units. In one trial, the subjects performed 150 pass-by interactions with the agents' active and passive strategies.

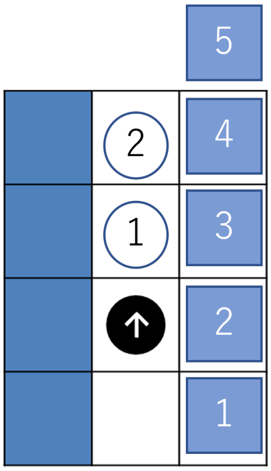

Figure 5. Ten types of relocation states with wall and agent combinations. The two circles are the positions of the two types of opponent agents and the five squares are the five types of the walls positions of narrow places.

In addition to the trials in which actions were measured, the same subjects were also asked to judge and record which strategy they thought the computer agents were using. The timing of agent relocation and strategy change was the same as in a previous trial. Three strategy estimation situations were recorded: an active strategy, a passive strategy, and when the subject felt that the agent was neither active nor passive.

Since the suitable action strategies changed in the same state depending on the action strategy of the opponent agent, we thought that we could observe changes in a subject's strategy by measuring states in which the subject's operations changed. Therefore, we prepared the following scores to determine when subjects changed their actions: a score of 1 when they changed their actions at the earliest stage, and a score of 0 when they chose actions only in response to the surroundings and regardless of an opponent agent's strategy.

The state in which a subject took two or more actions was marked as a state in which the strategy change was apparent, and the score was set to 1 when the first step after relocation was marked. As an example, Figure 6 shows a state transition in which a subject's action selections in Step 1 after relocation were multiple selections: a right turn 13 times and forward movement 17 times. Therefore, a state node after the relocation was marked, and the score of the state transition was set to 1. In the case of an unmarked state, the following transition was verified. Early and stable transitions were assumed to show more confidence in subjects' strategy changes in response to changes in agents' strategies. So, each step was multiplied by 0.8, and the weights corresponding to the number of transitions were multiplied. When there was no next step recorded, the score was set to 0.

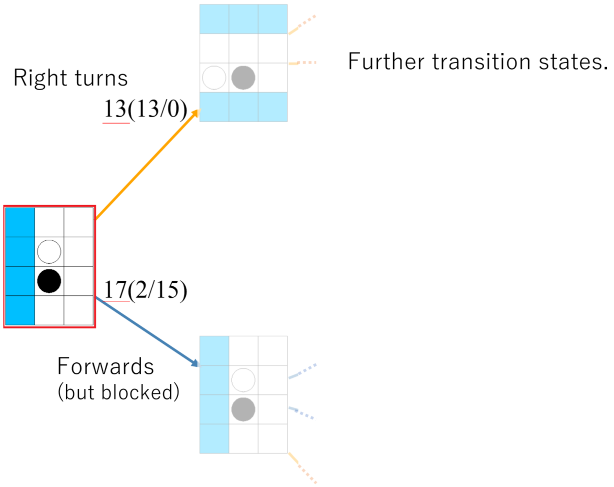

Figure 6. Example of score calculation.

4.3. Result

Figures 7 and 8 show the state transitions for Subjects 1 and 7, starting from where the subject agents were placed at the adjacent state up until the fifth step. From the left state, the steps transitioned to the right states one step at a time, following the arrows on the edges. The three numbers on the edges indicate the total number of transitions, the number of transitions when a computer agent had an active strategy, and the number of transitions when a computer agent had a passive strategy. If the agent's share of either strategy (active or passive) exceeded 80% of the total transitions, the edge was colored orange for the active strategy and blue for the passive strategy.

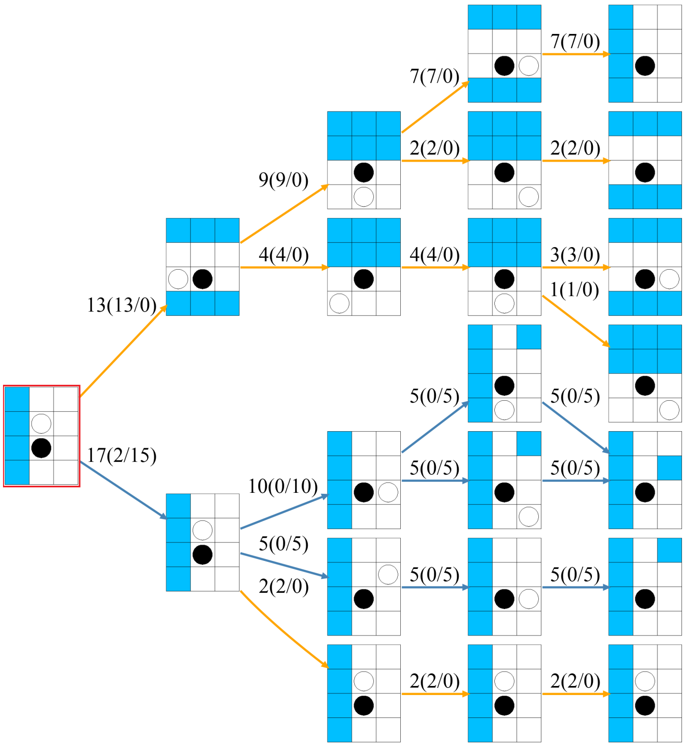

Figure 7. State transitions for Subject 1 starting from an adjacent state.

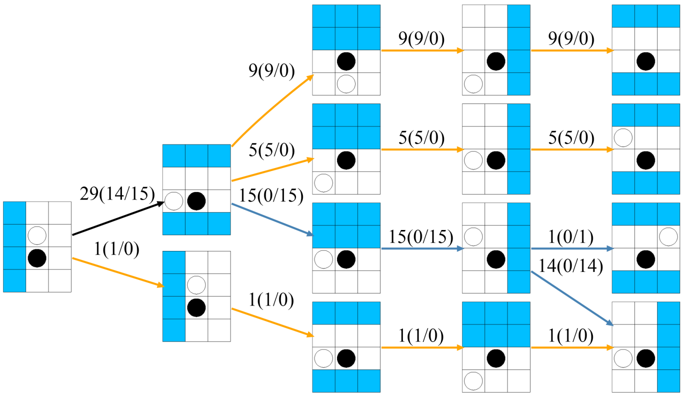

Figure 8. State transitions for Subject 7 starting from an adjacent state.

Table 1 shows the scores for each subject and each starting state for a second trial in which subjects were asked to indicate which strategy they thought the opponent agent was taking. State 2 in the table corresponds to the starting state in Figures 7 and 8. Other starting states—state 3, state 4, and state 8—are shown in Figure 9.

Table 1. Subjects' scores in the second trial.

	Subject 1	**Subject 2**	**Subject 3**	**Subject 4**	**Subject 5**	**Subject 6**	**Subject 7**
state 1	1.00	1.00	1.00	0.59	0.77	0.32	0.00
state 2	1.00	1.00	1.00	1.00	0.75	0.32	0.00
state 3	1.00	1.00	1.00	1.00	0.80	0.00	0.00
state 4	0.00	1.00	0.20	0.14	0.00	0.09	0.00
state 5	1.00	1.00	1.00	0.04	0.00	0.00	0.00
state 6	1.00	1.00	0.69	0.39	0.14	0.07	0.00
state 7	1.00	0.72	0.73	1.00	0.14	0.05	0.00
state 8	0.53	0.8	0.53	0.59	1.00	0.12	0.00
state 9	1.00	1.00	1.00	0.04	0.00	0.10	0.00
state 10	1.00	1.00	1.00	0.00	0.09	0.00	0.00
average	0.85	0.95	0.82	0.48	0.37	0.11	0.00
variance	0.11	0.01	0.08	0.18	0.16	0.01	0.00

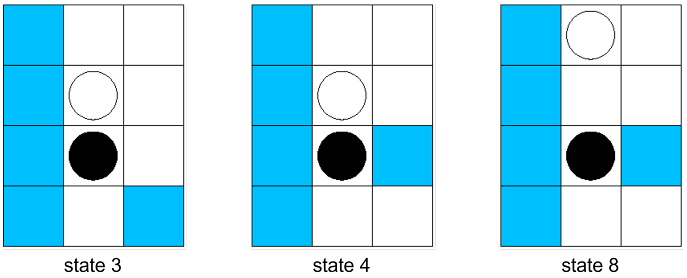

Figure 9. States 3, 4, and 8 in Table 1.

Figure 10 shows the strategy estimation status for each subject from the inputs of the second trial and an actual computer agent's strategy switching. The vertical axis represents the progress of relocation, the horizontal axis real represents strategies taken by the opponent agent, and 1 to 7 represent the agent's strategies as estimated by each subject.

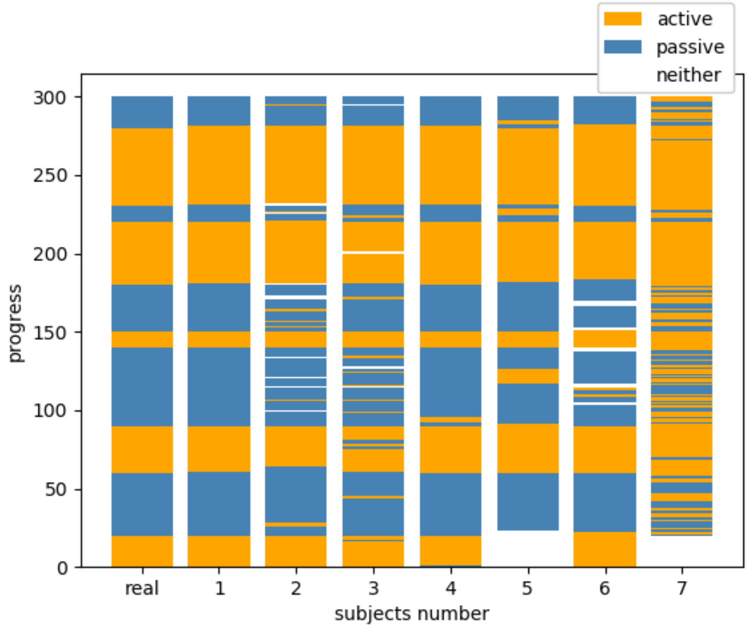

Figure 10. Computer agent's strategy switching and subjects' strategy estimation statuses in the second trial.

4.4. Discussion

As shown in Table 1, the scores for states 2 and 3 were close to 1 for subjects 1, 2, 3, 4, and 5, and they were close to 0 for subjects 6 and 7. In state 2, there was no narrow place in sight, and agents were adjacent. State 3 was after the subject agents crossed the narrow place and were adjacent to an opponent agent. Since these states were simple, with little influence from the surrounding terrain, it was thought that changes in action strategies were likely to occur according to each subject's Meta-Strategy.

In state 4, almost all subjects' scores were close to 0. State 4 was a situation in which the subject agent was in a narrow place, and a computer agent was blocking the exit. In this situation, even if the opponent agent had an active strategy, it gave way. Still, depending on whether it had an active or passive strategy, it moved either back in a straight line or sideways. As a result, the subject transitioned to a different state. Therefore, there was no opportunity to measure whether subjects would choose a different action in the same state, which would have resulted in a value close to 0 for the subjects.

The subject was also in a narrow place in state 8, one unit away from the computer agent. In this situation, the computer agent with a passive strategy moved forward, adjacent to the subject agent, and then tried to move to the side. If the computer and the subject agent tried to move forward simultaneously, the order of their actions was randomly determined, and if the computer agent acted first, it would be in state 4. Therefore, when starting from state 8, subjects lost a chance to measure their action change as if they had started from state 4 for about half of the time, depending on the random number of action orders.

Figures 11–17 show the scores for each subject by the time period, corresponding to the progress of the experiments. The periods were divided into the first 100 relocations, the second 200 relocations, and the entire second trial (300 relocations). In each period, the number of times that an opponent computer agent took two strategies was the same. For example, the values in columns 301 to 600 in Figure 14 are the scores in the column of Subject 4 in Table 1. Subject 4 scored 1.0 for the hree states 2, 3, and 7, so three points were plotted at the height of 1.0 in the right column of Figure 14.

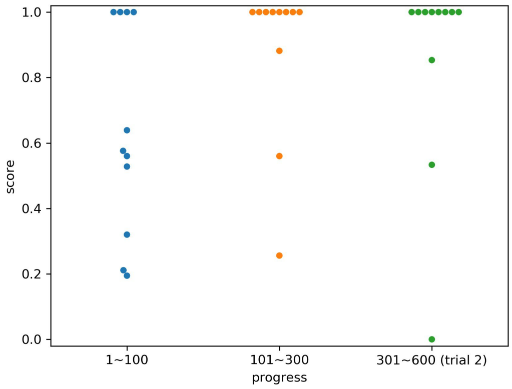

Figure 11. Scores of Subject 1 by time period.

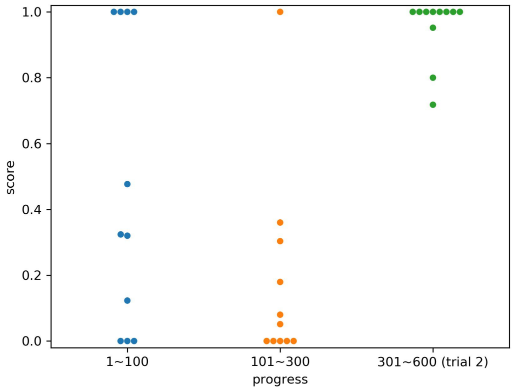

Figure 12. Scores of Subject 2 by time period.

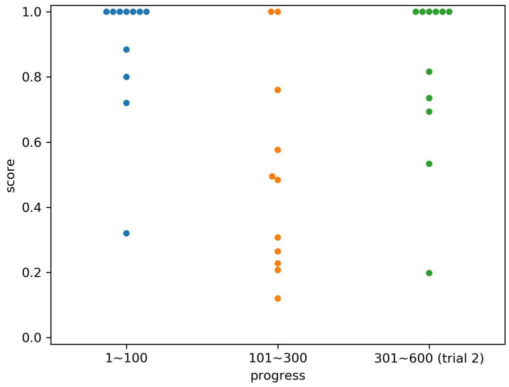

Figure 13. Scores of Subject 3 by time period.

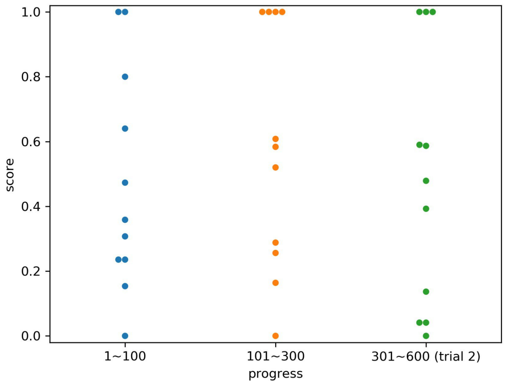

Figure 14. Scores of Subject 4 by time period.

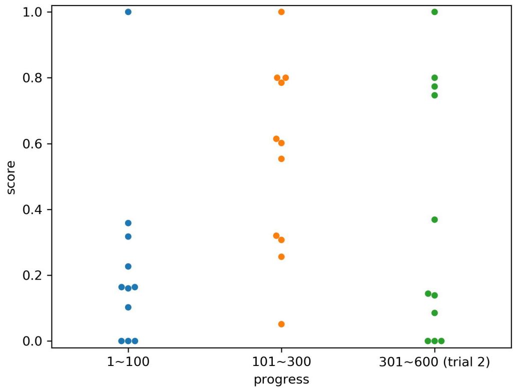

Figure 15. Scores of Subject 5 by time period.

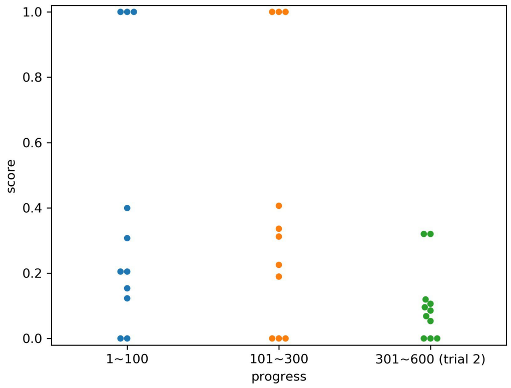

Figure 16. Scores of Subject 6 by time period.

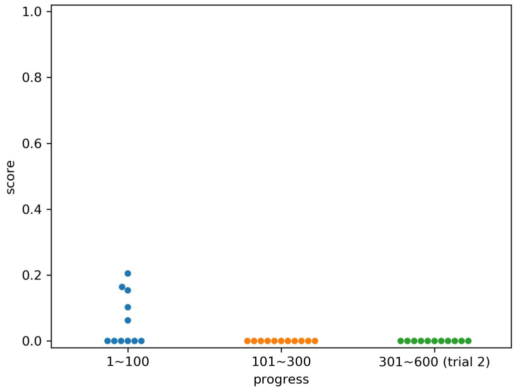

Figure 17. Scores of Subject 7 by time period.

Subjects 1, 2, and 3 had many state transitions with a score of 1. Thus, they chose their actions according to the strategy of the computer agent. This indicates that Subjects 1, 2, and 3 could estimate the computer agent's action strategy correctly.

Subjects 4 and 5 had a mixture of state transitions where the index score was close to 0 and 1. The initial states covered various patterns based on the spacing between agents and the locations of narrow places. In a non-symmetric state, there could be a bias regarding that agent that should change its action first. If a subject agent was placed in a position where it should move first, it was expected to change its action first, and its score would be 1—vice versa, its score would be 0. Even in this case, the subjects could predict the actions of the computer agent, so they could choose whether or not to change their actions. This means that the subjects were able to estimate the agent's strategies.

Subject 6 had a mix of scores that were close to 0 and 1 up to the trial's midpoint, but in the second trial, the scores shifted toward 0. It is possible that Subject 6, by interacting with an agent, decided not to consider the strategy, even though it could be estimated. Subject 2, conversely, changed to make choices in response to the opponent's strategy in the second half of the experiment.

Subject 7's score was biased toward 0. Subject 7's action choice did not change when the computer agent's strategy switched, suggesting that the subject could not estimate the computer agent's strategy.

Actually, in Figure 10, we can see that the strategy estimations of Subjects 1 and 4 were mainly correct. Subjects 2, 3, and 5 also made generally correct estimations. Subject 7's strategy estimation was incorrect in many moments, especially when an agent taking a passive strategy was mistakenly thought to be taking an active strategy. In particular, Figures 11 and 15 show that Subjects 1 and 7 had almost similar scores in the second half of the first trial (plotted in orange) and in the second trial (plotted in green). The action decisions and strategy estimations of the opponent agent were considered to have reached a certain degree at the time of the first trial. The estimation of the opponent's strategy was also expected to have been close to that in the second trial, which is shown in Figure 10.

Subjects 2, 3, 4, 5, and 6 showed differences in scores between the first and second trials. It is undeniable that the subjects were influenced by being taught that there were two strategies, active and passive, in order to assess the estimated situation, which may have caused a change in their estimations. With Subject 2, when the subject was able to choose actions according to the agent's two strategies after the instruction, it is possible that the recognition of the agent's two strategies led to the change in the action strategy.

Yokoyama et al. discussed pairs of action strategies [4], such as active and passive strategies, but did not appropriately evaluate Meta-Strategies for the selection of action strategies. This study tried to represent that point. We showed a tendency of Meta-Strategies toward the switching of action strategies during cooperative behavior. There are Meta-Strategies in which people always determine their strategies according to the results of their estimations of opponents' strategies or determine their strategies by considering which of them should move first depending on the state of the situation. Some Meta-Strategies determine the action strategies without being influenced by estimations of opponents' strategies, even though they are possible to estimate.

Furthermore, when considering a Meta-Strategy according to an opponent's action strategy, this Meta-Strategy is consistent in that it accords with the opponent's expressed strategy. Therefore, a Meta-Strategy can be considered passive in the Meta-Strategy layer even if it takes both active and passive action strategies. In contrast, a Meta-Strategy that does not involve the switching of action strategies is considered active.

5. Behavior of Agents according to Human Meta-Strategy Characteristics

We found that there are various types of human Meta-Strategies. Especially in the case of robots with a fixed primary user, such as one for household use, it is desirable to have them behave according to the characteristics of users' Meta-Strategies.

For users who are not affected by agents' behaviors, such as Subjects 6 and 7, it is not worthwhile for the robot to change its strategy. On the other hand, for users who can respond to changes in strategy, such as Subject 1, or users who can make choices depending on situations, such as Subjects 4 and 5, it is worth having robots switch strategies. We wanted to develop a method for calculating a score that can express the degree of strategy switching on a single axis by developing the score used in this study. For example, it will be possible to express the same degree as that of another user.

Furthermore, by calculating a new score that includes information on whether an opponent agent's strategy is active or passive after nodes where the subject has selected multiple actions, we expect to represent how subjects switch their strategies based on the estimated strategies of their opponent.

A model of an interaction strategy from a previous study [6] was considered, as shown in Figure 18. Individuals have a fixed personality layer as a base. On top of this, there is an emotional layer, including the dynamics of mental states during communication with others; this are expressed along axes such as interest and strain. Mental states change over time through interactions. It is expected that maintaining good mental states enables intelligent information processing (strategy selection) in communication in response to the approaches of others, as well as good interactions with agents.

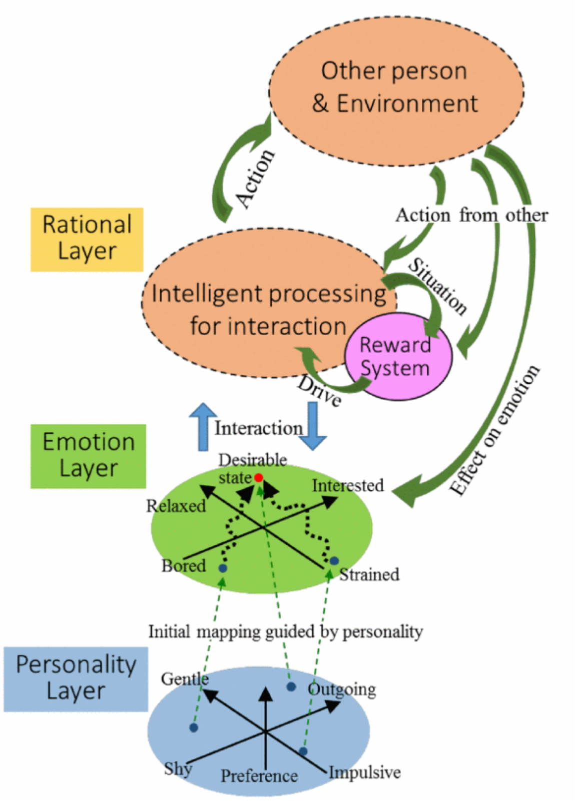

Figure 18. Model of emotion-based strategic interactions (Reprinted with permission from Ref. [6], 2015, T. Omori).

In this study, by measuring subjects' actions, it was confirmed that they used multiple strategies in communication as intelligent information processing, as shown in Figure 18. This is because the agents' strategies changed during interventions, and the subjects' strategies changed. The strategy switching of Subjects 2 and 6 changed drastically between the first and second halves of the experiment. This was not only due to a change in strategy selection caused by the change in the agent's strategies, but also due to the effects of mental states. Such a significant effect was not observed in the other subjects. The effects on mental states were different for each individual, even if the agent's behaviors were identical. We hypothesized that this might be due to the influence of personality, which is a lower layer. Therefore, we focused on the interpersonal control factor given by the revised scale of Machiavellianism (Mach) [17], in which each subject answered 28 questions on a five-point scale, ranging from strongly agree (five points) to completely disagree (one point). For each question, there was a weight for each characteristic, and the sum of the weights multiplied by the answers was the characteristic of a subject. The average of Mach for all subjects except Subject 5 was 20.092, with a standard deviation of 1.139. Thus, Subject 5, with 14.472, belonged to a different population from that of the other subjects, and it was considered to have different characteristics. However, unlike Subjects 2 and 6, the measured action choices did not change during the experiment. In addition, no significant differences were found in Mach between Subjects 2 and 6, who changed their choices of action strategies during the course of the experiment, and the other subjects, except for Subject 5 (Table 2). We could not determine the relationship between the interpersonal control factor measured by Mach and personality in the switching of strategies.

Table 2. Subjects' interpersonal control according to Machiavellianism.

Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Average	Variance
20.891	19.036	21.183	19.207	14.472	21.292	18.943	19.289	5.593

We would like to further examine which points affect mental states and how by including a method of calculating the scores in this study.

In the future, when robots are used to support people in their daily lives, while adapting to a primary user's personality, it is necessary to infer the user's mental states (e.g., a state of interest) and flexibly change action strategies accordingly.

6. Conclusions

The future goal of this research is to realize a robot that can predict its partner's action strategy by evaluating its partner's actions in comparison with its own action strategies. In this study, two different action strategies for the same task were learned by agents based on the idea that humans use multiple strategies when cooperating. We then conducted an experiment to evaluate the subjects' Meta-Strategies for using multiple action strategies with the agents.

We confirmed that the subjects estimated changes in the opponent agents' strategies, and then changed their strategies.

We also attempted to express when subjects' actions changed scores in response to agents with active and passive strategies. The goal was to evaluate differences in Meta-Strategies. We could represent the differences in Meta-Strategies that people have, albeit as a set of multiple scores. We found that there were several types of Meta-Strategies that people have. These different types can be considered active or passive in the Meta-Strategy layer.

In addition, we discussed a model of an interaction strategy. We confirmed the need for agents to flexibly make strategic choices according to people's Meta-Strategies and mental states. We would like to further study the behaviors of agents that have a positive influence on their counterparts.

Author Contributions: Methodology, K.M.; Software, K.M.; Supervision, N.W., O.N. and Y.T.; Writing—original draft, K.M.; Writing—review and editing, N.W., O.N. and Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, K.M., upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Itoda, K.; Watanabe, N.; Takefuji, Y. Analyzing human decision making process with intention estimation using cooperative pattern task. In Proceedings of the 10th International Conference on Artificial General Intelligence (AGI 2017), Melbourne, Australia, 15–18 August 2017; pp. 249–258.
2. Osawa, M.; Okuoka, K.; Sakamoto, T.; Ichikawa, J.; Imai, M. Other's Mind Model Based on Cognitive Interaction Framework. In Proceedings of the Human-Agent Interaction Symposium 2020, online, 7–8 March 2020; p. 40. (In Japanese)
3. Belpaeme, T.; Baxter, P.; Wood, R.; Cuayáhuitl, H.; Kiefer, B.; Racioppa, S.; Kruijff-Korbayová, I.; Athanasopoulos, G.; Enescu, V.; Looije, R.; et al. Mutlimodal child-robot interaction, Building social bonds. *J. Hum.-Robot. Interact.* **2013**, *1*, 33–53. [CrossRef]
4. Yokoyama, A.; Omori, T. Modeling of human intention estimation process in social interaction scene. In Proceedings of the 2010 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Barcelona, Spain, 18–23 July 2010; pp. 1–6.
5. Miyamoto, K.; Watanabe, N.; Takefuji, Y. Adaptation to Other Agent's Behavior Using Meta-Strategy Learning by Collision Avoidance Simulation. *Appl. Sci.* **2021**, 11, 1786. [CrossRef]
6. Omori, T.; Shimotomai, T.; Abe, K.; Nagai, T. Model of strategic behavior for interaction that guide others internal state. In Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Kobe, Japan, 31 August–4 September 2015; pp. 101–105. [CrossRef]

7. Abe, K.; Hamada, Y.; Nagai, T.; Shiomi, M.; Omori, T. Estimation of child personality for child-robot interaction. In Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Lisbon, Portugal, 28 August–1 September 2017; pp. 910–915. [CrossRef]
8. Hieida, C.; Abe, K.; Nagai, T.; Omori, T. Walking Hand-in-Hand Helps Relationship Building Between Child and Robot. *J. Robot. Mechatron.* **2020**, *32*, 8–20. [CrossRef]
9. Sugahara, R.; Katagami, D. Proposal of discommunication robot. In Proceedings of the First International Conference on Human-Agent Interaction, Sapporo, Japan, 7–9 August 2013.
10. Katagami, D.; Tanaka, Y. Change of impression resulting fromvoice in Discommunication motion of baby robot. In Proceedings of the HAI Symposium, Copenhagen, Denmark, 28–30 May 2015; pp. 171–176. (In Japanese)
11. Kozima, H.; Michalowski, M.P.; Nakagawa, C. Keepon: A playful robot for research, therapy, and entertainment. *Int. J. Soc. Robot.* **2009**, *1*, 3–18. [CrossRef]
12. Sato, T. Emergence of robust cooperative states by Iterative internalizations of opponents' personalized values in minority game. *J. Inf. Commun. Eng.* **2017**, *3*, 157–166.
13. Okamura, K.; Yamada, S. Empirical Evaluations of Framework for Adaptive Trust Calibration in Human-AI Cooperation. *IEEE Access* **2020**, *8*, 220335–220351. [CrossRef]
14. Yamada, K.; Takano, S.; Watanabe, S. Reinforcement Learning Approaches for Acquiring Conflict Avoidance Behaviors in Multi-Agent Systems. In Proceedings of the 2011 IEEE/SICE International Symposium on System Integration, Kyoto, Japan, 20–22 December 2011; pp. 679–684.
15. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
16. Abe, K.; Hieida, C.; Attamimi, M.; Nagai, T.; Shimotomai, T.; Omori, T.; Oka, N. Toward playmate robots that can play with children considering personality. In Proceedings of the Second International Conference on Human-Agent Interaction (HAI '14), New York, NY, USA, 29–31 October 2014; pp. 165–168.
17. Koga, H. A construction of revised machiavellianism scale. *Rikkyo Psychol. Res.* **2000**, *42*, 83–92. (In Japanese)