Effect of Robot and Screen Agent Recommendations on Human Decision-Making

Kazuhiko Shinozawa¹ and Junji Yamato²
¹ATR Intelligent Robotics and Communication Laboratories, ²NTT Communication Science Laboratories
Japan

1. Introduction

Two-dimensional (2D) character agents that have a human-like appearance are being developed. In the future, such agents will be able to interact with their users in a natural and friendly manner through speech recognition, synthesized speech, and action displays. In addition, robots or robotic companions that have a three-dimensional (3D) physical body are attracting attention as communication partners.

Such embodied social agents (ESAs) make interaction more meaningful than it is when interfaces do not appropriately display actions or speak (Beskow and McGlashan, 1997). It is known that people’s attitudes towards computerized media are similar to the attitudes they have towards other people (Reeves and Nass, 1996). Even if people only read text or hear a voice from computers, they tend to assign some social existence to them. More social richness, defined as more complete human-like presentations, promises to make computers more attractive, productive, and easy to use. Some research has provided fruitful results and suggestions for presentation, i.e., graphical appearance (Massaro, 1998), non-verbal behavior (Cassell and Thörisson, 1999), and speech characteristics (Nass and Gong, 1999), and for personality (Nass and Isbister, 1998), emotion (Ball and Breese, 1998; Becheiraz and Thalmann, 1998), ethnicity (Takeuchi et al., 1998), and interpersonal communication strategy (Shinozawa et al., 2001) as well. Much of such research suggests that an ESA should be an effective interface for interactions with media.

The above research mainly focused on graphical on-screen agents and computers. On the other hand, robots having a physical body have attracted some attention as useful physical agents, and the above research results may apply to interaction with such robots. However, when we consider robots as ESAs, a new research topic, “dimensionality”, appears. A robot has a 3D physical body while an on-screen agent has a 2D one. This leads to several questions: Does increasing dimension make a big difference or not? Does the physical 3D appearance affect us in a significant way during the interaction? When both a 2D agent like an on-screen agent and a physical 3D agent like a robot have a similar shape and use the same voice, what is the significance of the difference in dimensionality? Little research has focused on dimensionality, and we still have no solid answers. We live and work in 3D space. Everything has three dimensions and is located in 3D space. With a 3D body, pointing to some location makes it easy to understand what is being pointed at. When a
robot navigates a person, the combination of the robot's gestures and its body's direction has a strong relationship with high ratios of successful task completion (Ono et al., 2001). This suggests that the dimensionality produces a difference in the effect of interaction and that a 3D body makes interaction more meaningful.

How can we quantitatively measure the dimensionality or eye movement effects? Almost all research exploring the behavioral factor's effects of ESAs has been conducted by using questionnaire-based evaluations. For example, subjects are asked whether a robot is familiar or not. Getting a feeling of familiarity is important for a pet-like partner. However, it is not enough for a communication partner. Whenever we engage a communication partner, we listen to what the partner has to say, respond to it, and sometimes change our thinking. If an ESA's behavior can influence human decision-making, it leads that he/she treats it as a communication partner. This would be one of evaluations for a communication partner. So, we mainly investigate an ESA's influence on human decision-making for evaluating above factors.

In this chapter, in an attempt to answer some of the above questions, we discuss dimensionality, investigated by directly comparing results between an on-screen agent and a physical robot, and the role of a robot's eye movement in human-robot interaction. On both topics, the discussions are based on a quantitative evaluation of each factor's effect on human decision-making with a selection task. Section 2 describes the selection task with the direct comparison topic and presents the experimental results. Section 3 describes the effect of a tracking function in the 3D world case and presents the results. Section 4 discusses the effect of the dimensionality and eye movement with the experimental results. Then, Section 5 concludes with a short summary.

2. Difference in 3D and 2D agent’s recommendation

Recommendation in an advertisement and assistance in a navigation task are two typical situations influencing human decision-making. An ESA acting as an assistant can easily influence users' decisions because users want to know appropriate information. Generally speaking, however, changing a user's mind is difficult in the advertisement situation. Advertising is an important application of ESAs, and we can also say that the recommendation includes helpful interaction-like assistance, because the initial recommendation does not always depend on what the user wants. So, we focus on the advertisement situation and measure the influence of ESAs on human decision-making.

2.1 Color-name selection task

The color-name selection task was introduced to quantitatively measure the influence of ESAs on human decision-making in an advertisement situation (Shinozawa et al., 2001). In a color-name selection task, a subject looks at a colored region and selects the color name from two options. The matching ratio of the recommended color names is measured. The ratio is treated as showing the degree of a recommendation's influence on human decision-making.

2.2 Recommendation situation

In the case of direct comparison between an on-screen agent and a physical robot, two situations are considered. In one, an ESA points to an object located in 3D space and in the other, it points to an object in 2D space during interaction. Accordingly, we prepared two scenes for interacting with an ESA.
One scene is equivalent to the original one in the color-name selection task (Shinozawa et al., 2001). An ESA recommends a color name while it points to or looks at the color region and two color-name options on a CRT display when a subject should select one of them. In this case, objects used in the selection task are mainly located in 2D space. We therefore call this, the 2D world condition. The other scene is a new one that we call the 3D world condition. The color region that an ESA points to is in 3D space. Actually, we developed two new machines for displaying color regions in 3D space. One displays printed color plates according to external PC control. The other is a button box for displaying and selecting a color name. In the 3D world condition, a subject looks at printed color plates and selects color names using the button box (Fig. 1).

Figure 1. Display machines

The ESA recommended one of the two options under these two conditions, and we investigated the *dimensionality* by comparing the ESA’s effect on user decision-making.

### 2.3 Robot and On-screen agent

Appearance is important for robots as well as for on-screen agents. Humans tend to recognize social roles, gender, or characters by analogy with appearance. Prior knowledge according to appearance has much influence on subjective evaluation (Shibata and Tanie, 2001).

In *dimensionality* issue, to avoid such influence, we made the appearance of robots and on-screen agents as equivalent to each other as possible [Fig. 2(a) and (b)]. The robot’s height is 300 mm and on-screen agent’s height is similar to the robot’s one.

Figure 2. Appearance of ESAs

Similarly, voice plays an important role in molding the robot’s or on-screen agent’s character. Both the robot and on-screen agent use the same voice, which was made by
"Fluet", a Japanese speech synthesizer developed by NTT (Mizuno and Nakajima, 1998). The robot was also developed by NTT.

2.4 Gestures
We prepared 27 gestures for both the robot and on-screen agent, which included pointing to a color region, nodding, blinking, and so on. We made the robot’s motions similar to the on-screen agent’s motions.

2.5 Color names
Before starting an experiment, subjects were told that this was a color-name selection task and that they should make a selection based on their own feeling and that there were no correct answers. Most of the color regions and options for color names in the experiment, such as vermilion or carmine, are unfamiliar to ordinary people.

2.6 Speaking words
The robot and agent offered their personal opinions, for example, "I think it’s vermilion", "This shade is vermilion, isn’t it?", instead of making statements that would indicate it had definite knowledge about the displayed color. This was to avoid the effect of the subject’s attributing any authority to the robot and agent.

2.7 Recommended color names
We carried out pretests without an ESA’s recommendation and determined what color names the ESA should recommend and the orders of color name options. In both the 2D and 3D world condition, the same order of color names and the same recommended color name options were used.

2.8 Face direction when ESAs speak
Whenever we talk to someone, we look at that person’s face or eyes. ESAs should behave in the same way. With an on-screen 2D agent, a well-known illusion associated with full-faced portraits occurs: from any viewing angle, it appears that the agent’s gaze is always on the user (Bruce & Young, 1998). An on-screen agent’s full-face animation can give a feeling that the agent talks to subjects. However, humans can easily detect that the eyes of a 3D face’s are not looking at their face even if the difference from the correct direction is small. We therefore developed a subject’s head tracking function by which a robot adjusts its head direction so that its head faces subject’s head position.

2.9 Displayed Color region
In both world cases, the size of the displayed color region was about 270 mm x 160 mm, and the average distance from subjects to the color regions was about 600 mm. In the 2D world case, the colors displayed on the CRT were measured with the CRT color analyzer three times a day. The changes in these values were small (less than 10%) for the whole experiment. In the 3D world case, the colors displayed on printed plates were measured with a tristimulus colorimeter once a day. The changes in these values were also small (less than 7%) for the entire experiment. Therefore, all of the subjects saw the same color in each world case.
2.10 Subjects
Six experiments were conducted to manage all combinations described above.

1. 2D world condition
   (a) No recommendation (Group No2)
   (b) Agent recommendation (Group Ag2)
   (c) Robot recommendation (Group Ro2)

2. 3D world condition
   (a) No recommendation (Group No3)
   (b) Agent recommendation (Group Ag3)
   (c) Robot recommendation (Group Ro3)

None of the subjects were experts on color names and all were recruited from the general public. Each subject participated in only one experiment; never more than one. Table 1 shows information about the subjects in each group.

<table>
<thead>
<tr>
<th>Group</th>
<th>No2</th>
<th>Ag2</th>
<th>Ro2</th>
<th>No3</th>
<th>Ag3</th>
<th>Ro3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>31</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>Mean age</td>
<td>23.87</td>
<td>27.60</td>
<td>23.30</td>
<td>25.40</td>
<td>25.00</td>
<td>26.29</td>
</tr>
<tr>
<td>Max. age</td>
<td>49</td>
<td>39</td>
<td>36</td>
<td>36</td>
<td>39</td>
<td>45</td>
</tr>
<tr>
<td>Min. age</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Ratio of males</td>
<td>0.43</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.44</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 1. Subjects in each group

2.11 No recommendation case
To investigate the influence of an ESA’s recommendation on user decisions, we must know the mean of the matching ratios without recommendation. We therefore conducted no-recommendation experiments for the 2D and 3D world conditions. In these experiments, subjects did not see any on-screen agents or robots and chose a color name with no recommendation. In all recommendation conditions, the recommended color name options were fixed due to the pretest described above. The difference in matching ratios between the no-recommendation case and recommendation case shows the degree of the recommendation’s influence on user decision-making. When matching ratios in the recommendation case are greater than in the no-recommendation case, the influence is considered to be positive.

2.12 Procedure
Upon arriving at the lab, subjects were told that the purpose of this experiment was to mainly investigate the relationship between color regions and color names, and that they should make a selection based on their own feeling because there were no correct answers. After this explanation, they signed an informed consent statement.

2.12.1 No-recommendation
In the no-recommendation condition, subjects did the task without the ESA’s recommendation and there was no ESA near them.
2.12.2 Recommendation case
They encountered the robots or on-screen agents for the first time when they entered the experimental room. The ESA behaved like it was asleep until the subject pushed a button. Once the button had been pushed, the ESA behaved like it had been awakened and introduced itself, and the experiment started.

The experiment consisted of thirty questions, and each question had two possible responses. When presenting each question, the ESA made a statement endorsing one of the two possible responses. While the ESA was asking the question and presenting the two choices, these options appeared on the computer display in the 2D world condition and on the button box display in the 3D world condition. The subject in both conditions indicated his/her choice by clicking a radio button on the computer display and by pressing the corresponding button. The subject then pressed the "OK" button to send the selection to the computer. If the choice matched the ESA’s suggestion, the ESA nodded with approval while expressing a positive statement. If it did not match the suggestion, the ESA bowed and shook its head slowly while responding with a negative statement. This continued until all questions were answered. When the interaction finished, the experimenter gave the subject a questionnaire. Figures 3(a) and (b) show scenes of the on-screen agent and robot experiment in the 2D world condition, respectively.

Figure 3. Scene in 2D world case

Almost all subjects finished one experiment in less than 20 minutes. The options that subjects selected were automatically recorded in a computer when subjects pushed the OK button. The scenes in one experiment were videotaped.
2.13 Results
We calculated the mean matching ratios of the color names that the agent or robot successfully recommended to each subject. The mean ratios in the groups were also calculated. In the no-recommendation case, subjects did not get any recommendation, but the same color name options as in the recommendation case were presented. In the recommendation case, one of the color names was recommended. In estimating the mean matching ratios in the no-recommendation case, the mean matching ratios of the color names that were recommended in the recommendation case were calculated.

2.13.1 2D world case
Figure 5(a) shows the mean matching ratios for each group in the 2D world case. Factorial analyses of variance (ANOVA) were conducted for each mean of matching ratios. We compared the mean between no-recommendation, on-screen agent's recommendation, and robot's recommendation. There was a significant difference only between Group No2 and Group Ag2 (ANOVA, F=3.457, p=0.036, Scheffé, p=0.043). The difference between Group No2 and Group Ro2 was not statistically significant.

2.13.2 3D world case
Figure 5(b) shows the mean matching ratios for each group in the 3D world case. Again, factorial analyses of variance (ANOVA) were conducted for each mean of matching ratios. We again compared the mean between no-recommendation, on-screen agent's recommendation and robot's recommendation. There were significant differences between Group No3 and Group Ro3 and between Group Ag3 and Group Ro3 (ANOVA, F=6.725, p=0.002, Scheffé, p=0.003, p=0.042).

Table 2 summarizes the experiment results. The circles mean that the difference from the no-recommendation case is statistically significant.

<table>
<thead>
<tr>
<th></th>
<th>2D world case</th>
<th>3D world case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robot</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Effect of recommendation
3. Gaze effect in 3D agent and 3D world case

We used the robot with the subject’s head tracking function. In the actual experimental situation, subjects sat at fixed place and the relative position from a robot to the place was also fixed. Whenever the robot turned to the place with a fixed angle, its eyes could roughly catch the subject’s face or head. However, subjects’ sitting posture was not always same and their head sometime slightly moved during the experiment. Even in such a situation, does a robot need a tracking function? Is precisely facing a human head important for a 3D agent?

We conducted additional experiments to confirm the importance of having the robot directly facing subjects and investigated whether the importance also holds in a robot with a different appearance. The experiments consisted of three groups with same color-name selection task as in the above experiments. One group used the same rabbit-like robot without a tracking function. The robot could not adjust its head direction to the subject’s head movement and always made a same motion. For the other groups, a head robot made a recommendation instead of a rabbit-like robot. Figure 7 shows the head robot’s appearance. The robot was developed by MIT AI Laboratory and modified by NTT Communication Science Laboratories. The robot has only a head and neck (no arms or legs), both with 30 degrees of freedom. The robot can produce various facial expressions. In addition, each eyeball has a camera that can pan and tilt. Having cameras inside the eyeballs can makes the robot's gaze direction clear to subjects and ensures the center of the robot’s eye can be directed toward subjects with an appropriate vision system. To enable the tracking function, a skin-color region was detected using a detector developed by MIT AI Lab, and the eye direction turned to the center of that region. Table 3 summarizes the three conditions. Subjects’ mean age was 24.2 years.

![Figure 7. Head robot](image)

<table>
<thead>
<tr>
<th>Group</th>
<th>Ro3-2</th>
<th>Ro3-3</th>
<th>Ro3-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>30</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Mean age</td>
<td>20.97</td>
<td>25.07</td>
<td>23.21</td>
</tr>
<tr>
<td>Max. age</td>
<td>25</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>Min. age</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Ratio of males</td>
<td>0.5</td>
<td>0.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 3. Groups
3.1. Result in matching ratio
The mean of the matching ratio to the robot’s recommended options for each subject is shown in Fig. 8. Factorial ANOVA were conducted for each mean of matching ratios. We compared the mean between non-recommendation, without tracking, and with tracking for the same robot. There is a significant difference in the mean between non-recommendation and the robot with tracking for both robots. (In the rabbit robot case, ANOVA, F = 6.292, p = 0.003, Scheffé, p = 0.003 and in the head robot case, ANOVA, F = 4.759, p = 0.012, Scheffé, p = 0.015). The difference between the non-recommendation case and without tracking is not statistically significant (p > 0.1).

![Figure 8. Matching ratios](image1.png)

3.2. Result in mutual gaze
We determined a gaze-period according to the directions of both the robot’s gaze and subject’s gaze. Subjects’ gaze was classified into four categories: “color plate”, “button box”, “robot’s face”, and “other”. The robot’s gaze was classified into three categories: “color plate”, “subject’s face”, and “other”. This is because the robot could make its eyes turn toward the button box. In the head robot condition, we recorded the experiment scenes using the cameras equipped in head robot’s eyes, which gave us an accurate recording from

![Figure 9. Mutual gaze ratios](image2.png)
the robot’s viewpoint. All assignments were made on the basis of an experimenter’s observations with those videos. We assumed that a combination of a robot’s gaze to a subject’s face and a subject’s gaze to a robot’s face achieved mutual gaze. We defined the mutual gaze period as the period for which the combination continues. Figure 9 shows the mean ratio of mutual gaze periods for the whole experiment. The ratio in the tracking condition is significantly larger than in non-tracking condition (ANOVA, F=6.468, p=0.018).

4. Discussion

Figure 5 shows that the dimensionality of the ESA causes differences in the recommendation’s effect on user decision-making. The 3D body was not always superior to the 2D body for recommendation, and on-screen agents seem to have weak points, too. Those differences cannot be explained only by the advantages or disadvantages of pointing. In the 2D world condition, the color region was presented on the computer display. The display was in 3D space, so the color region was presented in the 3D space. From this point of view, there should be no difference between the 2D and 3D world conditions and therefore no difference in the effect. There must therefore be some other reasons. The results changed according to the combination of the location pointed to and the ESA’s dimensionality. This seems to be evidence supporting the importance of consistency in the dimensionality between communication partners and the environment consisting of the pointing location and manipulated objects. In the 2D world condition, the color region was displayed on the CRT and color name options that should be selected were also on the CRT. In the 2D world and on-screen agent case, all were contained in the frame of the computer display. The frame might have emphasized the appropriate consistency and the on-screen agent might have had a strong influence through its recommendation. In the 3D world condition, the communication environment was in the 3D physical space. In the 3D world and robot case, the color region and color name options were contained in the physical 3D space, although they were physically separated and there was no frame. In addition, the effect of robot’s recommendation was much greater than that of the on-screen agent’s. The robot’s body might have had a strong influence for emphasizing appropriate consistency to 3D space without a visual frame. In addition, behavior in communication contains ambiguous meanings and depends on the situation and communication environment (Sperber and Wilson, 1993). So, humans tend to quickly recognize an environment where the communication partner exists for interpretation. The results of our experiment seem to provide evidence that humans tend to quickly recognize communication environments even in interaction with an ESA, and also suggest that we must not forget the communication environment in designing ESA behavior.

4.1 Tracking function effect in 3D world case.

Figure 8 shows that the influence of a robot without a tracking function is still larger than both 2D agent’s and non-recommendation case, although the difference is not significant. The result suggests that the dimensionality still works even if a robot has no tracking function, although the effect is reduced. And, 3D shape with predetermined motions is not sufficient for significantly producing the robot’s dimensionality effect on human decision-making. In addition, those results also suggest that only a tracking function is not sufficient for
explaining all effects caused by the dimensionality, but that a tracking function must be important for robot’s influencing human decision-making. Much research has pointed out eye direction’s importance in conversation. For example, our gaze is one way of encouraging someone to talk (Michael & Mark, 1976). Gaze fixation exerts a special pressure to communication. In addition, research based on questionnaires has confirmed that a robot’s face direction makes a human notice its gaze (Imai et al, 2002) and suggested that its face direction is effective for signaling to whom the robot will talk in a multi-user situation.

As shown in Fig. 9, a tracking function increases robot’s mutual gaze chances when subjects may have a feeling that the robot looks at them. In addition, robots with a tracking function influenced decision-making more significantly than robots without it. Thus, such a feeling of being looked at would be necessary for producing some changes on human decision. In other words, our results show that mutual gaze influences not only human feelings but also decision-making in cases of interaction between humans and robots. And, a feeling that a robot looks at us is fundamental and crucial to a robot’s becoming a communication partner.

5. Conclusion

We experimentally confirmed through quantitative evaluation that the degree of recommendation effect firmly depends on the interaction environment. The results show that a three-dimensional body has some advantage when the interaction environment is a three-dimensional space, but has less advantage in two-dimensional space than a two-dimensional body does. This suggests that geometrical consistency between an ESA and the interaction environment plays an important role in communication.

In the 3D world case, we also experimentally confirmed that a tracking function for a robot can play the same important role that gaze has in humans; that is, it can increase the robot’s influence on human decision-making through the interaction even if the tracking movement is small.

For a robot as a physical advertisement agent, tracking is an important function. The tracking function successfully makes the mutual gaze ratio greater during interaction; however, the ratio doesn’t always have a relationship with the influence on decision-making.

Mutual gaze increases the chance when users may have feeling of being looked at, but it is not enough for insuring that a robot gives the feeling. To become a communication partner, a robot needs an effective method for emphasizing the feeling.

6. References


Bruce, V. & Young, A. (1998). In the eye of the beholder: The science of face perception: Oxford University Press.


Human-robot interaction research is diverse and covers a wide range of topics. All aspects of human factors and robotics are within the purview of HRI research so far as they provide insight into how to improve our understanding in developing effective tools, protocols, and systems to enhance HRI. For example, a significant research effort is being devoted to designing human-robot interface that makes it easier for the people to interact with robots. HRI is an extremely active research field where new and important work is being published at a fast pace. It is neither possible nor is it our intention to cover every important work in this important research field in one volume. However, we believe that HRI as a research field has matured enough to merit a compilation of the outstanding work in the field in the form of a book. This book, which presents outstanding work from the leading HRI researchers covering a wide spectrum of topics, is an effort to capture and present some of the important contributions in HRI in one volume. We hope that this book will benefit both experts and novice and provide a thorough understanding of the exciting field of HRI.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
