Using Balance Theory to Understand Social Agents

Hideyuki Nakanishi, Satoshi Nakazawa, and Toru Ishida
Department of Social Informatics, Kyoto University
Yoshida-Honmachi Sakyo-ku
Kyoto 606-8501, JAPAN
+81 75 753 4821
{nuka, nakazawa, ishida}@kuis.kyoto-u.ac.jp

Katsuya Takanashi
Communication Research Laboratory
2-2-2 Hikaridai Seika-cho Soraku-gun
Kyoto 619-0289, JAPAN
+81 774 95 2433
takanasi@crl.go.jp

1. Introduction

The graphical representation of an embodied conversational agent (ECA) depends on the style of its living world. It is hard to compare the appearance of a three-dimensional agent living in a virtual space with that of a flat agent moving on a Web browser. Similarly, it is also hard to compare the conversational ability of ECAs, since it depends on the agents' application domains. The conversation strategy of a tour-guide agent is different from that of a pedagogical agent.

We propose an alternative dimension to evaluate ECAs. That is how the agent can influence human relations. We focus on agents that enter human communities such as virtual worlds, not on the typical situation where an agent interacts with a single user [1]. We call these agents social agents. 'Social' is a key aspect additional to 'embodied' and 'conversational' in our method to evaluate agents. Currently, the social ability of agents is not as clear as the conversational ability and the quality of graphics. Since the criteria to evaluate the social aspect of agents are much ambiguous, we may have to understand social agents rather than evaluate them.

There is a study that tested an agent designed to support cross-cultural communication in a three-dimensional video chat world. In this study, it is found that the agent's behavior strongly influenced people's impressions of the agent, their conversation partners, and even stereotypes about their partner's nationality [6]. These results suggest that the agent plays a great role in human communities. Establishing relations with others is a basic aspect of sociality [2], and so an agent with sociality may influence human relations in the community. The evaluation of that influence is useful to develop good coordinator agents. And the result can complement graphical and conversational evaluations. At least, the high degree of the influence indicates that the agent is not ignored by people.

We propose the balance theory as a tool to evaluate the agent's influence on human relations. The balance theory states that two people's relations depend on whether both persons have the same sentiment toward a certain object [4]. If this theory can be successfully applied to the relationships between the agent and two people, we can see the agent's influence on the relationship between two people. To confirm this idea, we investigated whether our agent prototype can play the role of such an object based on this theory. We tested the capability of the agent to win a favorable feeling from both people or from only one side while the other side develops an unfavorable feeling towards the agent. And we tried to observe people's relations change according to the balance theory.

2. Balance Theory

Since the balance theory can explain interaction in human relations [8], we verified that similar interaction could

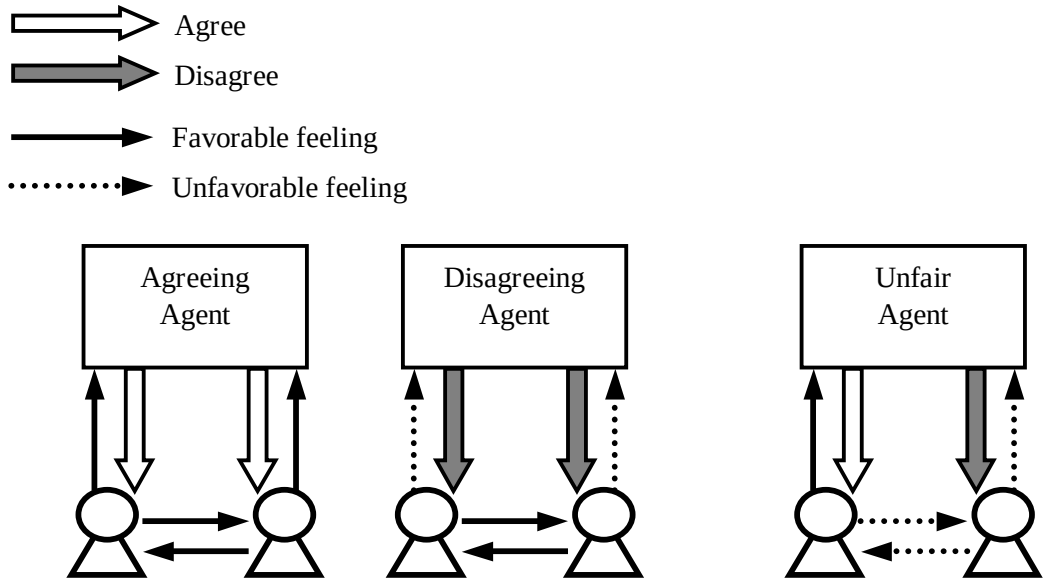
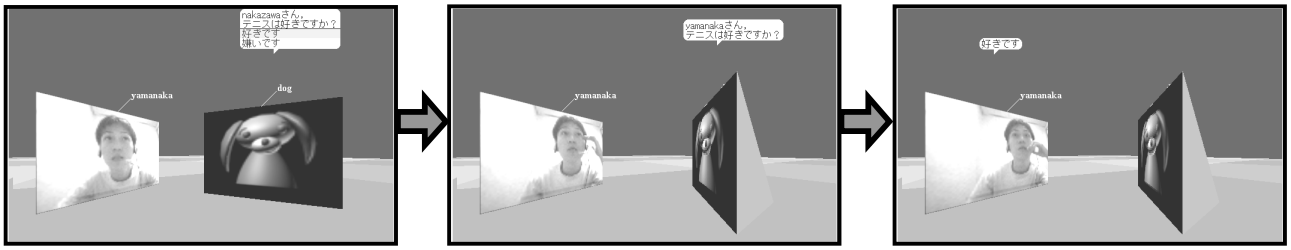


Figure 1. Balance theory with an agent and two humans

occur between a human-agent relation and human relations. The balance theory can be applied to the relations between two people and an object X, which can be a person, a thing, or a fact [5, 7]. When you have a positive or negative sentiment toward X and think that your partner has the same sentiment toward X, you have a positive sentiment toward your partner. If you think that your partner has a different sentiment toward X, you have a negative sentiment toward your partner. If X is a person, this theory explains the human relations among three people. In our experiment, X is an agent. We observed a situation in which the agent tries to control the sentiment of the relation between two subjects by controlling each subject's sentiment toward the agent. Actually, X can be a thing or a fact, but we found that a fact only has enough influence to cause agreement or disagreement when it involves a controversial issue or a strong like/dislike in food, and that in any case it cannot change the direction of agreement/disagreement. However, the agent itself differs from an issue or food, since the agent autonomously establishes its relations with others by communicating in the same way people do. It is difficult to match two people's sentiments to things or facts by controlling one's sentiment, but this may be possible when an agent is used as an object in this way.

We made the agent express an agreeing or disagreeing attitude to observe how subjects develop favorable/unfavorable feelings toward the agent. This would be verified if a subject had more positive sentiment toward the agent when its attitude is agreement than when its attitude is disagreement. We assume that subjects think their partners would respond to the agent similarly. On the assumption that the agent can control its impression, we tried to determine whether the balance theory works in the following three conditions. The first condition is that the agreeing agent agrees to both subjects' opinions. The second is that the disagreeing agent disagrees with both of their opinions. The third is that the unfair agent agrees to one subject's opinion but disagrees with the other subject's opinion at once. Figure 1 shows the three situations where the balance theory is valid. The theory is valid if a subject comes to have more positive sentiment toward his/her partner when the agreeing/disagreeing agent shows an agreeing/disagreeing attitude to both subjects than when the unfair agent shows a different attitude for each subject.

The agent's influence may be weakened when human communication channels widen. We tried to confirm that the agent's influence is weaker when the two subjects have a conversation than when they do not. Additionally, we compared an environment where the content of the conversation between the agent and the partner is hidden



(a) (b) (c)

Figure 2. Simultaneous environments. (a) You answer the agent. (b) The agent asks your partner. (c) Your partner answers the agent.

with another environment where their conversation is known. We aim to determine whether the subjects' sentiments toward the agent influence their relations. However, we expect their relations may inversely influence their sentiments toward the agent if their communication channel is wide.

3. Statistical Experiment

We conducted a statistical experiment to test our evaluation method. In a first-time meeting of about fifteen to twenty minutes, one subject met another subject and our agent to interact in the three-dimensional video chat environment FreeWalk [10]. In this meeting, the subject established relations with his/her partner as well as the agent. In the experiment, before beginning the meetings, subjects were only told that they would interact with another subject and the agent. The agent acted completely automatically without any intelligent functionality. It had static sentences for talking and a prepared script to guide its behavior. Voice and live video connections were available for subjects to communicate with one another, but they had to communicate with the agent through menu-based interaction. A menu, which contained a question from the agent, and menu items, which were the choices the subjects could use to answer to the question, were presented inside a balloon displayed on the subjects' computer screens. The agent asked a subject a question while showing a menu, and then the subject clicked a button in the menu to answer the question. Figure 2 briefly illustrates this interaction.

The reason for using the three-dimensional video chat system was to compare four distinct environments formed by the effects of two factors, each of which has two levels. One factor is the steps of establishing agent-human and human relations, which are sequential or simultaneous. In the sequential environments, subjects develop their sentiments toward the agent before they develop their sentiment toward their partner by learning their partner's sentiments toward the agent. In the simultaneous environments, these steps are done concurrently. The other factor is whether subjects can talk with their partners or not. We prepared conversational environments and non-conversational environments. We assumed that both of the two factors would widen or narrow the communication channel between the two subjects.

In the simultaneous conversational and non-conversational environments, two subjects and the agent formed a circle to have a conversation (see Figure 2). The agent provided both subjects with twenty questions asking their likes and dislikes about various things. Based on the questionnaire about likes and dislikes, which provided a priori data, twenty questions were chosen as follows. For half of the questions, the answers of the two subjects matched, and their answers to the other half of the questions did not match. To show the agent's attitude, it responded to the subjects' answers as follows. When the two subjects chose the same answer, the agreeing agent agreed with the answer, the disagreeing agent disagreed with it, and the unfair agent showed a neutral attitude. When their chosen answers did not match, the agreeing and disagreeing agents showed a neutral attitude, but the unfair agent agreed to one answer and disagreed to another answer. In the simultaneous conversational

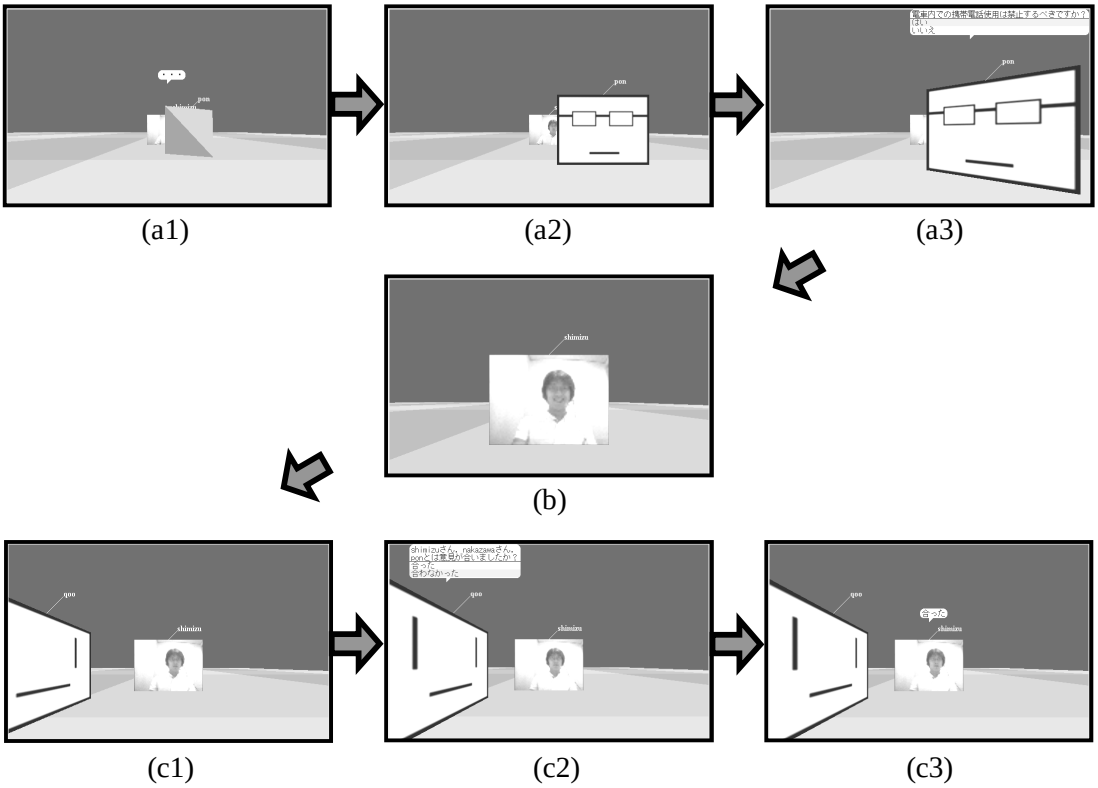


Figure 3. Sequential environments. (a1) After the agent talks to your partner (you cannot read their utterances), (a2) it approaches you (a3) to ask you. (b) You talk with your partner. (c1) The second agent appears (c2) to ask both of you, (c3) and then both answer.

environment, the agent interrupted subjects' ongoing conversation to ask questions. In the simultaneous non-conversational environment, subjects answered the agents' questions without having any conversation with their partners.

In the sequential conversational and non-conversational environments, the agent traveled between the two subjects standing far from each other in the virtual space to repeat a one-to-one conversation with each subject (see the three pictures, (a1), (a2), and (a3) in Figure 3). During the conversation between the agent and a subject, his/her partner could not read the text in the balloons on the display. In a phase of the agent's behavior, it asked the subject four questions about social problems and then asked the same four questions to the other subject. The agent repeated this phase three times to provide a total of twelve questions. The agreeing agent agreed with both subjects for ten questions, but disagreed on the other two questions. Similarly, the other two agents showed their intended attitudes for ten questions and the opposite attitudes for the rest. We mixed two exceptions with ten intended responses because in the preliminary experiment we found that a few exceptions made the intended attitude clearer. After the agent finished asking questions and left the subjects, a second agent arrived to ask subjects eight questions about favorable feelings toward the agent. The three pictures ((c1), (c2), and (c3)) in Figure 3 show this scene. The face of the second agent was different from the first agent's face. The second agent had a conversation with both subjects together so that each subject would know the partner's sentiment toward the agent. In the sequential conversational environment, subjects approached to talk with each other during an interval between each phase of the agent's behavior. A meeting had two intervals and the period of an interval was a minute, so subjects had a conversation for two minutes. Figure 3(b) shows the scene of this conversation.

All subjects were university students. A total of 185 people (113 male and 72 female) participated in our

experiment. After the meetings, subjects answered the questionnaire about the agent, the agent from the partner's point of view, and the partner in terms of similarity and attractiveness. As the result of analyzing the questionnaire data, we found the agent could influence human relations under the situation in which people could not have a conversation, and the agent established relations with them before they established their relations, while each subject could not know what the agent and the other subject was talking about. In the case that agent-human relations and human relations were established simultaneously, the agent's influence on human relations became a little weaker. Even in the case when the agent established relations with people beforehand, the agent lost his ability to influence their relations, if they were provided with a few chances to talk with one another during the establishment process. If the agent had to join the conversation of people to try to establish relations with them and to influence their relations, it was very hard for the agent to influence their sentiments toward it, which is the preliminary step in influencing their relations.

4. Follow-up Analysis of Conversation

Why do agents lose the power of influence for human relations when subjects talk to each other? In order to find the cause, conversation recorded in the simultaneous conversational environment was examined by conversation analysis [9]. Within this environment, agent-human conversation is carried out through the text channel, and human-human conversation is carried out through the vocal-speech channel. But unlike the statistical experiment, a WOZ (Wizard of Oz) agent, which is controlled by the experimenter secretly, is introduced to make its behaviors look more natural, by presenting topics and giving inductive talk before questions.

(Ex1) is a transcription of a part of the conversation. In the transcription, A and B are subjects, and X is the agent. 'A=>B' means A speaks to B. Italic sentences are talk through text channel. Some exchanges of subjects through the vocal-speech channel (=>2) are inserted between the two parts of the adjacent pair on the text channel, the question of agent (=>1) and the subject's answer (=>3). This phenomenon shows conversation between subjects often includes remarks about the agent-human exchange itself.

(Ex 1) Remarks about the agent-human exchange
=>1 X=>A: *Mr. A, this is very off the subject though, do you often listen to music?*
=>2 A=>B: That is very off the subject, don't you think? (laugh)
=>2 B=>A: I think so, too. Who would have expected a dog to have consideration? (laugh) It seems that he is paying attention to the flow of conversation.
=>2 A=>B: He is clever, in a sense. (laugh)
A=>B: …Well, often.
=>3 A=>X: *Yes*
X=>A: *I see.*

The vocal-speech channel between subjects is used as another communication channel that is placed at the meta-level in agent-human exchange. Through this channel, subjects often evaluate the behaviors of the agent and reach an agreement. The underlying assumption made here by both subjects is that the agent cannot understand their conversation through this channel. Therefore, could the influence of the agent be sustained by limiting the occurrence of this parallel channel? To examine this point, an experiment using text chat was carried out. As a result, occurrences of a parallel channel decrease drastically. However, the result of questionnaire was not so much different from the one in the simultaneous conversational environment of FreeWalk. There are many unsolved problems that cannot be resolved even by restraining the parallel channel. As a serious problem, disagreement expressed by the agent (=>1) causes the antipathy of the subjects toward it (=>2) and leads to sympathy between them in (Ex 2).

(Ex 2) Sympathy between subjects

=>1 <X> *I do not hit it off well with Mr. A, because you want to visit Universal Studio Japan.*

<A>…*Fine.*

=>2 <B> *Well, I think Mr. X is kind of rude.*

<A> *I'm afraid I'll never get along with him.*

One of the reasons why such sympathy is caused is the form of disagreement utterance of the agent. In general, while agreement can be expressed directly and immediately, disagreement utterances are accompanied by many devices, such as hesitation, indirect and mitigated expression, giving a reason for the disagreement, and so on [9]. In this regard, the disagreement utterance of the agent, which lacks these devices, sounds unnaturally strong and therefore rude. Another reason is subjects' consideration of the 'face' of participants. People in public places generally take care to "save face," and they also take similar measures for the 'face' of others [3]. For example, when someone stumbles over a stone on the road, people around him tend to pretend not to notice it. Similarly, in (Ex2), subject B willingly tries to recover the partner's face after it is threatened by explicit disagreement of the agent in the public conversation (=>2), and this motivates the subjects to have sympathy for each other. The agent's infelicitous behaviors caused by its inability to engage in natural conversation can lead to the antipathy of the subjects toward the agent and sympathy with each other. As a result, this reduces the effect on human relations that the agent has aimed for.

5. Conclusion

We conducted an experiment to test our evaluation method to apply the balance theory to the relationship between social agents and people. We investigated whether the balance theory could be applied to explain the interaction among agent-human relations and human relations. The result of statistical and conversation analyses demonstrate that the evaluation of the agent's influence on human relations is useful to assess the agent's ability to control its own social behavior appropriately as well as its conversational ability.

References

1. Bickmore, T. and Cassell, J., Relational Agents: A Model and Implementation of Building User Trust, *CHI-2001*, pp. 396-403, 2001.
2. Damon, W., *Social and Personality Development: Infancy Through Adolescence*, W.W. Norton & Company, 1983.
3. Goffman, E., *The Presentation of Self in Everyday Life*, Doubleday & Company Inc., 1959.
4. Heider, F., *The Psychology of Interpersonal Relations*, Wiley, 1958.
5. Horowitz, W.M., Lyons, J. and Perlmutter, H.V., Induction of Forces in Discussion Groups, *Human Relations*, Vol. 4, pp. 57-76, 1951.
6. Isbister, K., Nakanishi, H., Ishida T. and Nass, C., Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space, *CHI-2000*, pp. 57-64, 2000.
7. Jordan, N., Behavioral Force That Are A Function of Attitudes and of Cognitive Organization, *Human Relations*, Vol. 6, pp. 273-287, 1953.
8. Kogan, N. and Tagiuri, R., Interpersonal Preference and Cognitive Organization, *Journal of Abnormal and Social Psychology*, Vol. 56, pp. 113-116, 1958.
9. Levinson, S.C., *Pragmatics*, Cambridge University Press, 1983.
10. Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T., FreeWalk: A 3D Virtual Space for Casual Meetings, *IEEE MultiMedia*, Vol. 6, No. 2, pp. 20-28, 1999.