Analysis of Human-Human and Human-Computer Agent Interactions from the Viewpoint of Design of and Attribution to a Partner

Kazuhisa Miwa (miwa@is.nagoya-u.ac.jp)
Hitoshi Terai (terai@cog.human.nagoya-u.ac.jp)
Graduate School of Information Science, Nagoya University
Nagoya, 464-8601 JAPAN

Abstract

Recently, not only Human-Human Interaction (HHI) but also Human-Agent Interaction (HAI) where humans and cognitive artifacts such as computer agents collaborate have emerged. To investigate the nature of such interaction it is important to deal with two perspectives separately: design of and attribution to computer agents. The perspective of attribution is what a human attributes to a computer agent whereas the perspective of design is how a computer agent is actually designed. We propose the "illusion experiment paradigm" where we can control independently these two factors. Two experiments were performed in which a pair of subjects solved a simple reasoning task collaboratively. We analyzed how their hypothesis formation and testing behavior were influenced by these two factors. Experimental results basically indicated that subject problem solving behavior was only influenced by the factor of design, whereas their reciprocity behavior as one representative social behavior was influenced by both design and attribution factors.

Keywords: Problem solving,; Collaboration; Interaction.

Introduction

We generate new values through interactions with external environments while revising, processing, and generating information. Recently, not only Human-Human Interaction (HHI) but also Human-Agent Interaction (HAI) where humans and cognitive artifacts such as computer agents collaborate have emerged; many researchers have begun to show interest in the nature of such interactions.

One research paradigm emerging at the intersection of the HHI and HAI studies is the Media Equation framework (Reeves & Nass, 1996). A finding obtained in the framework is that human beings often relate to computer or television programs as they relate to other human beings. This view has also begun to provide new principles for designing computer agents (Dryer, 1999).

An important suggestion that surfaced from Media Equation studies is that when studying HAI it is important to deal with two perspectives separately: *design* of and *attribution* to computer agents. The perspective of *attribution* includes what a human attributes to a computer agent: i.e., he/she recognizes a partner as a human or another artificial agent whereas the perspective of *design* is how a computer agent is designed: i.e., to what degree the agent is actually constructed as sophisticatedly as humans behave.

Some interesting findings have indicated the importance of the perspectives of *attribution* and *design*, including such a traditional example as ELIZA where humans converse with a simple computer program called ELIZA as they talk with humans, even though the program generates only very superficial responses (Weizenbaum, 1966). In the issue of the uncanny valley, pointed out in android science where the appearance of robots is very closely designed to humans, familiarity to robots decreases rather extremely, and the uncanny valley emerges (see a CogSci2005 workshop site: http://www.androidscience.com/). These results imply that in an investigation of the nature of HAI, two factors, *design* of and *attribution* to computer agents, should be dealt with separately.

In the preceding studies, our two different factors, *attribution* and *design*, were manipulated either together or a single factor was dealt with. In this study, we propose the "illusion experiment paradigm" where we control them independently. The objective of this study is to clarify the nature of HHI and HAI based on these two crucial factors: *design* of and *attribution* to computer agents.

Additionally, the experiments conducted in the Media Equation paradigm have mainly focused on understanding the human social relationship with computer agents by measuring subjects' impressions of the agents with questionnaires. In this study we measure subject problem solving behavior as a dependent variable while controlling *attribution* and *design* as independent variables. By measuring the nature of interactions emerging in behavior that are objectively observable, such as problem solving performance in addition to subjective estimations, we expect to discuss the nature of HHI and HAI based on more established empirical evidence.

Task

2-4-6 task

In this study, we use Wason's 2-4-6 task as an experimental task (Wason, 1960) because it has been used as a standard experimental task in studies on human discovery, and its nature is well understood (Newstead & Evans, 1995). The standard procedure of the 2-4-6 task is as follows. All subjects are required to find a rule of a relationship among three numerals. The rule that subjects find is called a target rule. In the most popular situation, a set of three numerals, "2, 4, 6," is presented to subjects in the initial stage. The subjects form a "hypothesis" about the regularity of the numerals based on the presented set. An example hypothesis is "three continuous even numbers." The subjects then produce a new set of three numerals and present it to the experimenter. This set is called an instance. The experimenter gives a Yes as feedback to the subjects if the set produced by the subjects is an instance of the target rule, or a No as feedback if it is not an instance of the target rule. The subjects continuously carry out experiments, receive feedback from each experiment, and search for the target.

Important concepts

First, we briefly explain important concepts regarding the two key factors: i.e., the nature of the targets that the subjects try to find and the hypothesis-testing employed by the subjects.

The nature of targets: We categorize the targets from the viewpoint of their generality. We define targets as broad targets if the proportion of their members (positive instances) to all instances (all sets of three numerals) in the search space is large. On the other hand, we define targets as narrow targets if the same proportion is small. An example of the former type of target is "the product of three numerals is even" (where the proportion of target instances to all possible instances is 7/8), and an example of the latter type is "three evens" (where the proportion is 1/8).

Hypothesis testing: There are two types of hypothesis testing: a positive test and a negative test. The positive test (P-test) is conducted in an instance where the subject expects there to be a target. That is, the P-test is a hypothesis test using a positive instance for a hypothesis. The negative test (N-test) is, in contrast, a hypothesis test using a negative instance for a hypothesis. For example, if a hypothesis were about "ascending numbers," the P-test would use a sequence like "1, 3, 9"; the N-test would use a sequence like "1, 5, 2."

Illusion Experiment Paradigm

In this study, we control *design* of and *attribution* to a partner agent independently by developing the illusion experiment paradigm (see Figure 1).

Pairs of subjects separated into different rooms participated in the experiment. Each subject sat in front of a computer terminal through which he/she solved the 2-4-6 task collaboratively with a partner (see Figure 2). Each subject could refer to the partner's hypothesis. Until the experiment's end, they were permitted to generate twenty instances to identify the target rule. That is, they observed a total of twenty-one instances including the first one, "2, 4, 6," indicated by the system. Each of the two subjects alternately generated instances, thus each generated ten of the twenty instances. Each subject could refer to instances generated by the partner.

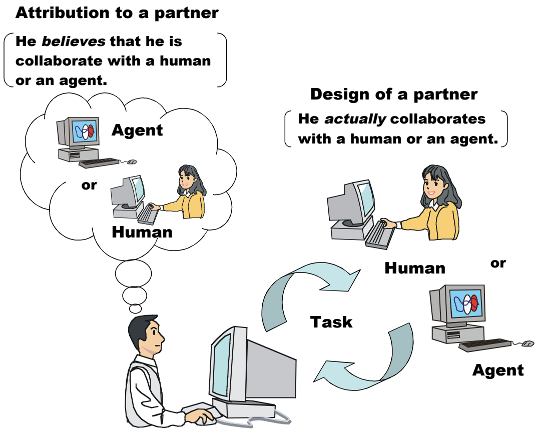
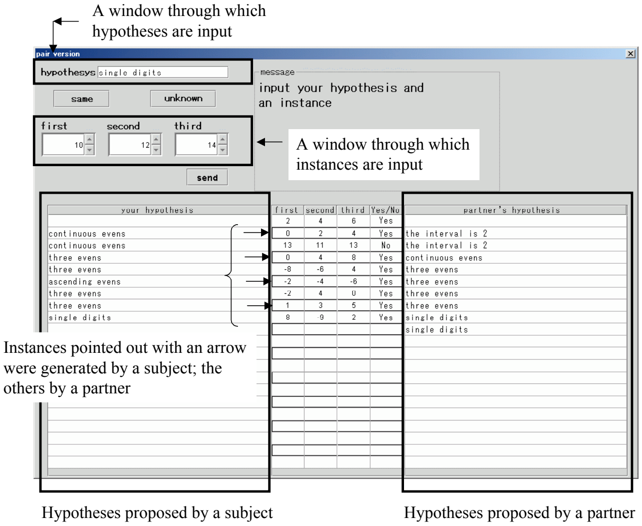

Attribution to a partner
He *believes* that he is collaborate with a human or an agent.
Agent
or
Human
Design of a partner
He *actually* collaborates with a human or an agent.
Human
or
Task
Agent

Figure 1: Two factors, design of and attribution to a partner, in the illusion experiment paradigm

A window through which hypotheses are input



pair version

hypothesys: single digits

same | unknown

message

input your hypothesis and an instance

first: 10 second: 12 third: 14

send

A window through which instances are input

your hypothesis	first	second	third	Yes/No	partner's hypothesis
	2	4	6	Yes	
continuous evens	0	2	4	Yes	the interval is 2
continuous evens	13	11	13	No	the interval is 2
three evens	0	4	8	Yes	continuous evens
three evens	-8	-6	4	Yes	three evens
ascending evens	-2	-4	-6	Yes	three evens
three evens	-2	4	0	Yes	three evens
three evens	1	3	5	Yes	three evens
single digits	8	-9	2	Yes	single digits
					single digits

Instances pointed out with an arrow were generated by a subject; the others by a partner

Hypotheses proposed by a subject

Hypotheses proposed by a partner

Figure 2: Example screenshot of the experimental environment

Design of a partner: The first experimental factor as an independent variable was related to the *design* of a partner agent. This factor was controlled by manipulating with which partner subjects actually collaborated. Three cases were set up: (1) collaboration with a human subject (w/ Human), and (2) collaboration with a computer agent. The former represents a case where a partner computer agent was sophisticatedly designed where it behaves almost the same as humans. The latter case was subdivided into two sub cases: (2a) collaboration with an agent who uses the positive test strategy in hypothesis testing (w/ P-test Agent), and (2b) collaboration with an agent who uses the negative test strategy (w/ N-test Agent). The reason for adopting these strategies as the factor of *design* is that this issue has been recognized as one of the most important topics in the human discovery process (Klayman & Ha, 1987; Laughlin, et al., 1987).

Attribution to a partner: The second factor was related to *attribution* to a partner brought about by the experimenter's instructions. This factor was controlled by manipulating with which subjects they believed to be collaborating. Two cases were set up: (1) a case where subjects were instructed to collaborate with a program installed on a computer they were manipulating, and (2) a case where they were to collaborate with a human subject in a different room, communicating by the Internet.

The first factor (*design*) was manipulated as follows. When collaborating with a human subject, each terminal was connected to the Internet by wireless LAN, and each subject solved the problem with a partner in a different room through the Internet. On the other hand, when collaborating with a computer agent, each terminal operated independently from the others and each subject solved the task with an agent established on a computer. The agent, i.e., the computational problem solver, was developed in the author's previous study (Miwa, 2004).

The second factor (*attribution*) was controlled according to the experimenter's instructions. When leading subjects into a collaboration situation with a human subject, a terminal was connected to an Internet socket with a dummy cable, and subjects were deceived into believing interaction with a partner in a different room. On the other hand, when collaboration with a computer agent was instructed, the dummy cable was removed; subjects thought that their terminal worked independently because the Internet connection was achieved by wireless LAN.

Experiment 1

Experimental design

In Experiment 1, we conducted analysis on subject problem solving behavior by adopting the following two dependent variables, hypothesis testing and hypothesis formation, because these two types of behavior have been crucial throughout the history of the laboratory studies of human discovery processes (e.g., Gorman, 1992; Klahr, 2000).

Each subject discovered two kinds of target rules. One target was "the product is 48," and the other was "three different numbers." The former is an example of a narrow target and the latter is a broad target. The order of the targets used in the experiment was counter-balanced.

This was a three (*design*) x two (*attribution*) design with both *design* (Human, P-test Agent, and N-test Agent) and *attribution* (human and agent) as between-subject variables. A total of ninety-six undergraduates participated in the experiment, and they were assigned to one of the six experimental conditions as evenly as possible.

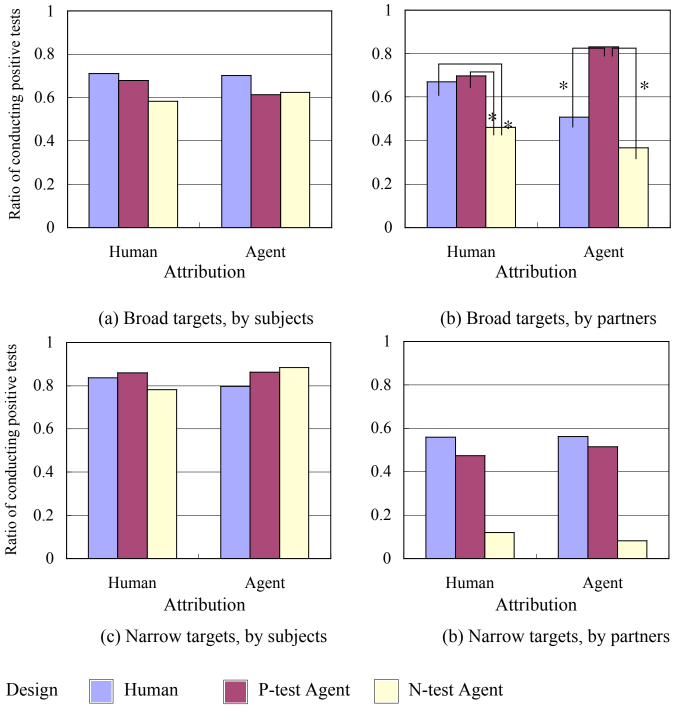
Results

Hypothesis testing

Here we discuss how subject hypothesis testing strategies as a dependent variable is influenced by two independent variables, *design* of and *attribution* to a partner. Cognitive psychological studies on human hypothesis testing have indicated that humans have a strong bias for conducting positive tests rather than negative tests (Mahoney & DeMonbruen, 1997; Mynatt, et al., 1977). This bias is called the positive test bias. To what degree does this bias change in each type of collaboration dealt with in this study?

Figure 3 shows the ratio of positive instances for subjects' hypotheses, separated into instances generated by the subjects themselves and instances by their partners. In other words, Figs. 3(a) and (c) show the ratio of conducting the positive test in the subjects' hypothesis testing, while Figs. 3(b) and (d) show the ratio of their partners' instances fulfilling the positive test for subjects' hypothesis.

Figures 3(a) and (c) show that the ratio of positive tests in subject hypothesis testing was invariable regardless of the change of partners. A two (*attribution*) x three (*design*) ANOVA did not reveal any significance (in finding the broad target the main effect of *attribution*: F < 1; the main effect of *design*: F(2, 90)=1.59, p > 0.1; the interaction: F < 1, in finding the narrow target the main effect of *attribution*: F < 1; the main effect of *design*: F < 1; the interaction: F < 1).



(a) Broad targets, by subjects

(b) Broad targets, by partners

(c) Narrow targets, by subjects

(b) Narrow targets, by partners

Figure 3: Ratio of conducting positive tests

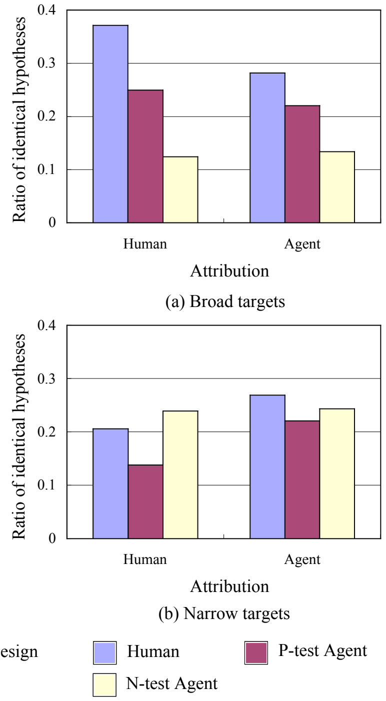
This point becomes more interesting when we compare collaboration with the P-test Agent and the N-test Agent, where partner's hypothesis testing strategy was controlled. Figures 3(b) and (d) show that in collaboration with the P-test Agent, the ratio of partner's instances fulfilling the positive test for subjects' hypothesis was higher than in collaboration with the N-test Agent. A two (*attribution*) x three (*design*) ANOVA revealed that the interaction between the two factors was significant in Fig. 3(b) (F(2, 90) = 4.21, $p < 0.05$), and a significant difference by LDS analysis is indicated by a "*" in the figure (MSe=0.0455, $p < 0.05$). The same ANOVA reveals that the main effect of *design* was significant in Fig. 3(d) (F(2, 90) = 35.87, $p < 0.01$), and LDS analysis indicated that the ratios in the Human and P-test Agent conditions were higher than in the N-test Agent condition (MSe=0.0537, $p < 0.05$). Neither the main effect of *attribution* nor the interaction was significant ($F < 1$, $F < 1$, respectively).

This means that even though the quality of information of the instances given by a partner varied depending on the change of a partner's hypothesis testing strategy (*design* of a partner), this did not influence the subjects' positive test bias. Moreover, this consistency did not depend on the experimenter's instructions as to whether the subjects collaborated with a human subject or with a computer agent (*attribution* to a partner).

Hypothesis formation

Laughlin & Futoran (1985) indicated that in group activities an individual accepts another group members' hypothesis as his/her own hypothesis by accurately estimating the validity of others' hypotheses accurately, which creates the superiority of group activities to individual activities. Next, we discuss how subjects' references to their partner's hypothesis in hypothesis formation is influenced by changing a partner.

Figure 4 shows the ratio of cases in which subjects proposed an identical hypothesis to the partner's when they revised their own hypothesis. A two (*attribution*) x three (*design*) ANOVA revealed that the main effect of *design* was significant in finding the broad target (F(2, 90) = 7.71, $p < 0.01$), and an LSD test showed that ratios in the Human and P-test Agent conditions were higher than in the N-test Agent condition (MSe=0.0407, $p < 0.05$). Neither the main effect of *attribution* nor the interaction was significant ($F < 1$, $F < 1$, respectively). In finding the narrow target, no statistically significant effect was found (the main effect of *attribution*: F(1, 90) = 1.50, $p > 0.1$; the main effect *design*: $F < 1$; the interaction: $F < 1$). This means that in such cases subjects' tendency to adjust their hypothesis to their partner's hypothesis became stronger when collaboration was with the P-test Agent compared to collaboration with the N-test Agent (*design* of a partner). This tendency did not depend on the experimenter's instruction as to whether the subjects collaborated with a human subject or with a computer agent (*attribution* to a partner).



(a) Broad targets

(b) Narrow targets

Figure 4: Ratio of adjustment of subjects' hypothesis to partner's

Experiment 2

Reciprocity behavior

The results of Experiment 1 showed that the factor of *attribution* to a partner did not influence subject problem solving behavior whereas the factor of *design* of a partner did in some cases. The question in Experiment 2 is whether similar results to Experiment 1, where problem solving behavior was analyzed, are obtained in subject social behavior. In Experiment 2, we dealt with reciprocity behavior, a most representative human social behavior, and investigated how reciprocity behavior is influenced by *design* of and *attribution* to a partner. Reciprocity behavior is behavior where people help those who help them. This behavior is a strong universal norm across all human cultures.

To analyze reciprocity behavior, in Experiment 2 the following experimental procedures were added to Experiment 1. In Experiment 1, all hypotheses formed by subjects were presented to a partner. However in Experiment 2 the subjects were asked to select whether they presented their hypotheses to a partner. In the initial stage, the subjects were given 500 Japanese-yen as their monetary resource. It costs 25 yen to present a hypothesis to a partner. Throughout the experiment, a total of twenty hypotheses is formed. Therefore, if all hypotheses are presented to a partner, all 500 yen is spent; they lose their money.

The subjects were instructed that their partner also performed the task under the same conditions. When the partner's hypothesis was not presented, a cell of the computer screen that indicated the partner's hypothesis was masked with a gray tile. Additionally, on the screen, the subject's and partner's remaining money is indicated throughout the experiment.

In this situation, the subjects were required to solve two tasks continuously. In the first task they were presented with 80% of the hypotheses from the partner; in the second task they were presented with only 20%. Therefore, in the first task the partner is left with only 100 yen, but 400 yen in the second task at the final stage of the experiment. We analyzed how the subjects' frequency of presenting hypotheses to a partner varies when the ratio of the partner's presentation of hypotheses decreases.

Experimental design

Each subject found two kinds of target rules in the two tasks explained above. Both target rules were broad: one was "three different numbers," and the other was "the product of three numbers is even." The order of the targets used in the experiment was counter balanced.

This was a mixed two (*design*) x two (*attribution*) x two (ratio of the partner's hypothesis presentation) design, with *design* (P-test and N-test agents) as a between-subject variable, *attribution* (human and agent) as a between-subject variable, and ratio (80% and 20%) as a within-subject variable. In Experiment 1, the subjects collaborated with a human partner, however in Experiment 2, the subjects only collaborated with a computer agent excluding the condition, where they collaborated with a human, in the factor of *design*. A total of eighty-nine undergraduates participated in the experiment, and they were assigned to one of the four experimental conditions as evenly as possible.

Results

Figure 5 shows the ratio of the frequencies of subjects' presenting hypotheses to a partner in each of the experimental conditions. Figure 5 shows that when instruction was collaboration with a human the ratio of presenting hypotheses to a partner decreased from 80% of the partner's hypotheses was received to only 20% was received; when the instruction was collaboration with a computer agent the ratio did not vary. This result indicates that reciprocity behavior was influenced by the factor of *attribution* to a partner. A three (*attribution*) x two (*design*) x two (ratio of the partner's hypotheses presentation) ANOVA revealed that the interaction among the three factors was not significant (F < 1). The same ANOVA revealed that the interaction between *attribution* and ratio of hypotheses presentation was significant (F(1, 85) = 19.39, p < 0.01). The simple main effect of the ratio of hypothesis presentation was significant in collaboration with a human (F(1, 85) = 46.45, p < 0.01) but not in collaboration with a computer agent (F < 1).

The ANOVA also revealed that the interaction between *design* and ratio of hypotheses presentation was significant (F(1, 85) = 6.65, p < 0.05). The simple main effect of the ratio of hypothesis presentation was significant in collaboration with the N-test agent (F(1, 85) = 30.53, p < 0.01) but not in collaboration with the P-test agent (F(1, 85) = 3.53, n.s.). This means that reciprocity behavior was also influenced by the factor of *design* of a partner.

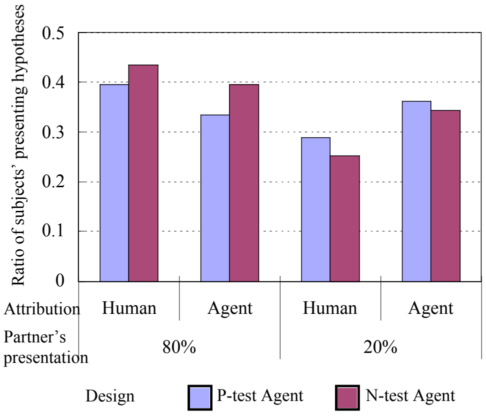


Figure 5: Ratio of subjects' presenting hypotheses to a partner

Discussion and conclusion

Attribution to a partner

In Experiment 2, subject problem solving behavior such as hypothesis formation and testing was also analyzed to confirm the results of Experiment 1. Table 1 shows the overall results of Experiments 1 and 2 indicating how two factors, *attribution* and *design*, influenced subject problem solving and social behavior.

Media Equation studies have indicated much empirical evidence that humans provide similar interactions with computer agents as interaction with humans in various aspects of human cognitive and social activities (e.g., Fogg & Nass, 1997, Moon & Nass, 1996, Nass, et. al., 1994; 1995; 1999). In our study, this finding corresponds to a situation where the subjects' behavior does not vary between situations in which they believe they are collaborating with a human and with a computer agent; i.e., the subjects' behavior is not influenced by the factor of *attribution* to a partner.

The results indicated in Table 1 show that subjects' behavior was more greatly influenced by the factor of *design* of a partner than by the factor of *attribution* to a partner. This result is consistent with the findings of the Media Equation studies. However, note that an effect of the factor of *attribution* emerged in highly socialized interaction such as reciprocity behavior. This point should be carefully investigated in future works to discuss the generality of the results obtained in the Media Equation studies.

Table 1: Overall results of Experiments 1 and 2

		design	attribution
Exp. 1	testing	No	No
	formation	Yes*	No
Exp. 2	testing	Yes	No
	formation	Yes	No
	reciprocity	Yes	Yes

*: only when finding the general target

Design of a partner

Subject hypothesis formation behavior, such as the degree of adjusting hypotheses to the partner's, was strongly influenced by the partner's hypothesis testing strategy as the factor of *design* of a partner when finding the broad targets. This evidence was very strong throughout Experiments 1 and 2. In cases of collaboration with a computer agent, the algorithm in the agent's hypothesis formation was consistent; therefore subjects were presented with similar hypotheses under the P-test Agent and N-test Agent experimental conditions. However, it is interesting that subjects tended to adjust their hypothesis to the partner's more remarkably when only collaborating with the P-test Agent.

In finding the broad target, the possibility of the agent receiving a Yes as feedback in the experiment was much higher than in finding the narrow target (see the definition of types of target). Actually the former possibility was 0.75, whereas the latter was 0.21 in Experiment 1. Therefore, the P-test Agent faced many positive hits by receiving a Yes feedback repeatedly, confirming its hypothesis many times (Klayman & Ha, 1987; Miwa, 2004). On the other hand, the N-test agent faced negative hits falsifying its hypothesis. From the viewpoint of the subjects who observed the agent's activity, this means that the P-test Agent seems to propose a reliable hypothesis whereas the N-test Agent usually proposed a dubious one. This difference produced the result that in finding the broad target subjects tended to adjust their hypothesis to the P-test agent's hypothesis.

References

Dryer, D. C. (1999). Getting personal with computers: How to design personalities for agents. Applied Artificial Intelligence, 13, 273-295.

Fogg, B. J., & Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. International Journal of Human-Computer Studies, 46, 551-561.

Gorman, M. (1992). Simulating science: heuristics, mental models, and technoscientific thinking. Indiana university press.

Klahr, D. (2000). Exploring science: The cognition and development of discovery processes. Cambridge, Mass.: MIT Press.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review, 94, 211-228.

Laughlin, P. R., & Futoran, G. C. (1985). Collective induction: Social combination and sequential transition. Journal of Personality and Social Psychology, 48, 608-613.

Laughlin, P. R., Magley, V. J., & Shupe, E. I. (1997). Positive and Negative Hypothesis Testing by Cooperative Groups. Organizational Behavior and human decision processes, 69, 265-275.

Mahoney, M. J., & DeMonbruen, B. G. (1997). Psychology of the scientist: an analysis of problem solving bias. Cognitive Therapy and Research, 1, 229-238.

Miwa, K. (2004). Collaborative discovery in a simple reasoning task. Cognitive Systems Research, 5, 41-62.

Moon, Y. & Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. Communication Research, 23, 651-674.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. Quarterly Journal of Experimental Psychology, 24, 326-329.

Nass, C., Steuer, J., Henriksen, L., and Dryer, D. C. (1994). Machines, social attributions, and ethopoeia: performance assessments of computers subsequent to “self-” or “other-” evaluations. International Journal of Human Computer Studies, 40, 543-559.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, D. C. (1995). Can computer personalities be human personalities? International Journal of Human Computer Studies, 43, 223-239.

Nass, C., Moon, Y., and Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. Journal of Applied Social Psychology, 29, 1093-1110.

Newstead, S., & Evans, J. (Eds.). (1995). Perspectives on Thinking and Reasoning. UK: Lawrence Erlbaum Associates Ltd.

Reeves, B., & Nass, C. (1996). The Media Equation: how people treat computers, television, and new media like real people and places. CSLI Publications.

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly journal of experimental psychology, 12, 129-140.

Weizenbaum, J. (1996). A computer program for the study of natural language communication between man and machine. Communications of the Association for Computing Machinery, 9, 36-45.