An Experimental Study on Collaborative Scientific Activities with an Actual/Imaginary Partner

Kazuhisa Miwa, Yoshie Baba, and Hitoshi Terai
Graduate School of Information Science
Nagoya University
{miwa, baba, terai}@cog.human.nagoya-u.ac.jp

Abstract. In this study, we experimentally investigate collaborative scientific activities that are undertaken through a virtual space such as the Internet. In such cases, a partner has two aspects: an *imaginary partner* with whom the problem solver seems to work together, and an *actual partner* with whom he/she actually works. We design an experimental environment in which we can control the two factors independently. The experimental result shows: (1) a bias appearing in human behavior, such as the positive test bias in hypothesis testing, was not influenced by the change of an *actual partner*; however (2) the degree of using information given by a partner, such as reference to a partner's hypothesis, varied considerably with the change of an *actual partner*. Neither phenomenon above depended on the type of *imaginary partner*.

Keywords: Scientific discovery, Collaborative problem solving, Internet, Hypothesis testing

INTRODUCTION

In this study, we experimentally investigate collaborative scientific activities that are undertaken through a virtual space such as the Internet. In particular, we focus on hypothesis formation and verification stages, which are the most typical processes in scientific activities. In cognitive science, there is an enormous number of studies on scientific discovery by problem solvers working solo or in pairs (e.g., Gorman, 1992; Klahr, 2000). On the other hand, there are only a few studies on collaborative scientific discovery in non-face-to-face situations, which are engaged via virtual space.

A partner (collaborator) is an important factor in determining collaborative scientific activities. In the case of collaboration via virtual space, a partner has two aspects: an *imaginary partner* with whom the problem solver seems to work together, and an *actual partner* with whom he/she actually works; additionally, the situation may arise where the two aspects of a partner are not identical.

Recently, we have begun to experience not only HHI (Human-Human Interaction) but also HAI (Human-Agent Interaction), and the interest in such interaction is increasing. Reeves & Nass (1996) proposed the Media Equation framework in which they concluded that human beings relate to computer or television programs in the same way they relate to other human beings. Following after their framework, many studies investigated how computational agents act effectively like human instructors (e.g., Baylor, 2000; Moreno, et al., 2001; Morishita et al., 2004). In such interaction, there may be a case where people believe they are collaborating with humans (*imaginary partners*) but are actually collaborating with computational agents (*actual partners*). The reverse relation is also possible. Additionally, we can intentionally set up the above situation such as designing an environment for CSCL. We believe that it is important to study this type of interaction not only for responding to pragmatic requirements in designing CSCL and CSCW environments, but also for understanding fundamental features on human collaborative problem solving.

In this study, we design an experimental environment where we can control experimental factors on both *imaginary* and *actual* partners independently, and we investigate the above issues, focusing on hypothesis formation and testing in scientific activities. Concretely, in each of the various collaboration situations, which were constructed with combinations of *imaginary* and *actual* partners, we attempt to understand (1) through hypothesis verification how people's hypothesis-testing strategy is influenced by a partner's strategies, and (2) via hypothesis formation how people change to refer to another hypothesis given by a partner.

TASK AND BACKGROUND

2-4-6 task

In this study, we use Wason's 2-4-6 task as an experimental task (Wason, 1960). The reason for using this task is that it has been used as a standard experimental task in studies on human discovery, and that the nature of the task is well understood (Newstead & Evans, 1995). The standard procedure of the 2-4-6 task is as follows (see Table 1). All subjects are required to find a rule of a relationship among three numerals. In Table 1, the target rule is "three evens." In the most popular situation, a set of three numerals, "2, 4, 6," is presented to subjects at the initial stage. The subjects form a "Hypothesis" about the regularity of the numerals based on the presented set. 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 to find the target.

Table 1 Example process of solving the 2-4-6 task.

Hypothesis	Instance	Feedback	States
	2, 4, 6	Yes	
Three continuous evens	6, 8, 10	Yes	Positive hits
Three continuous evens	0, 10, 30	Yes	Negative hits
The interval is the same	0, 10, -5	No	False negatives
The interval is the same	0, 10, 50	Yes	Negative hits
Ascending numbers	1, 5, 100	No	False positives
...			

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."

Klayman & Ha (1987) summarized states in which a subject's hypothesis is falsified (see the "States" column in Table 1). Let us consider a case where the target is "three evens" and the subject's hypothesis is "ascending numbers." When the subject conducts a P-test using the instance "1, 3, 5" and then receives a No feedback, his/her hypothesis is disconfirmed (*false positives*). Another state of conclusive falsification is caused by the combination of a N-test and a Yes feedback, using the instance "8, 6, 2" (*negative hits*). On the other hand, states of ambiguous verification are obtained from the combination of a P-test and a Yes feedback, using "4, 6, 8" (*positive hits*), or the combination of a N-test and a No feedback, using "5, 3, 1" (*false negatives*).

EXPERIMENTAL DESIGN

Design

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. Each subject could refer to the partner's hypothesis. Until the end of an experiment 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. Figure 1 shows an example screenshot of a terminal in the process of the experiment.

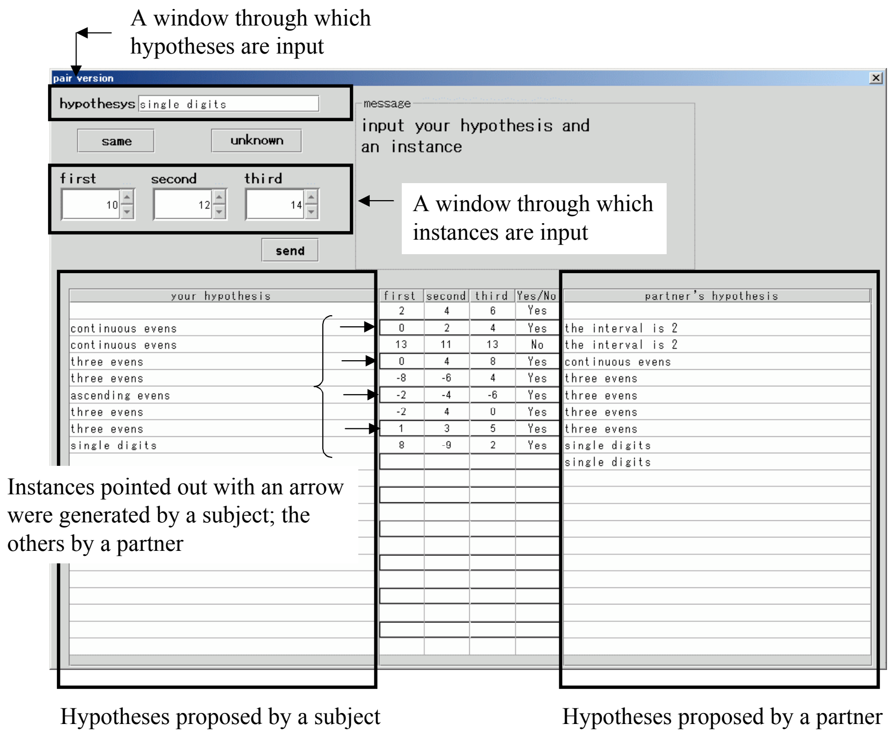

Figure 1 An example screenshot of the experimental environment.

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

Experimental factors

A three (*actual partners*) x two (*imaginary partners* brought about by the experimenter's instruction) between-subjects design experiment was conducted. Table 2 shows the numbers of subjects assigned to each condition. A total of ninety-six undergraduates participated in the experiment.

Table 2 Numbers of subjects participating in the experiment.

		Imaginary partners (Instruction)	
		Human	Agent
Actual partners	Human	16	16
	Positive-test Agent	15	17
	Negative-test Agent	16	16

The first factor was related to an *actual partner*, where three cases were set up: (1) a case of collaboration with a human subject (w/ Human), and (2) a case of collaboration with a computer agent. 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 in this study 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).

The second factor was related to an *imaginary partner* brought about by the experimenter's instruction. 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, with whom they could communicate via the Internet.

Method of controlling the factors

The first factor (an *actual partner*) was manipulated as follows. When collaborating with a human subject, each terminal was connected to the Internet via wireless LAN, and each subject solved the problem with a partner in a different room via the Internet. On the other hand, in the case of 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 preceding study (Miwa, 2004).

The second factor (an *imaginary partner*) was controlled according to the experimenter's instruction. When leading the subjects into a situation of collaboration with an imaginary human subject, a terminal was connected to an Internet socket with a dummy cable, and the subjects were guided to imagine interaction with a partner in a different room. On the other hand, when collaborating with an imaginary computer agent, the dummy cable was removed; the subjects imagined that their terminal worked independently because the connection with the Internet was achieved via wireless LAN.

RESULTS

Subjective estimation of the goodness of collaboration

After the main experimental session, the subjects were required to estimate the degree of goodness of collaboration with a partner on a scale of 1 to 5. Figure 2 shows the average of the estimated degree in each experimental condition. A 2 (*imaginary partners*) x 3 (*actual partners*) ANOVA revealed that the main effect of an *actual partner* reached significance ($F(2, 90)=10.79$, $p < 0.01$); a LSD analysis showed that the degree with Human was higher than that with the P-test Agent and the degree with the P-test Agent was higher than that with the N-test Agent ($MSe=0.8874$, $p < 0.05$). On the other hand, the main effect of an *imaginary partner* was only marginally significant ($F(1, 90)=3.29$, $p < 0.1$). The interaction of the two factors was not significant ($F < 1$).

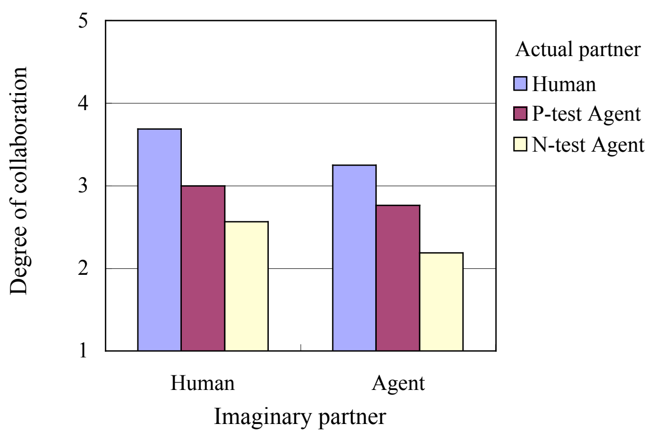

Figure 2 Degree of subjective estimation on goodness of collaboration.

Hypothesis testing

Here we discuss how subjects' hypothesis testing strategies are influenced by a change of 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 instances being positive ones for subjects' hypotheses, separated into instances generated by 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 the subjects' hypothesis.

Figures 3(a) and (c) show that the ratio of the positive test in the subjects' hypothesis testing was invariable regardless of the change of partners. A 2 (*imaginary partners*) x 3 (*actual partners*) ANOVA did not reveal any significance (in finding the broad target the main effect of an *imaginary partner*: $F < 1$; the main effect of an *actual partner*: $F(2, 90)=1.59$, $p > 0.1$; the interaction: $F < 1$, in finding the narrow target the main effect of an *imaginary partner*: $F < 1$; the main effect of an *actual partner*: $F < 1$; the interaction: $F < 1$).

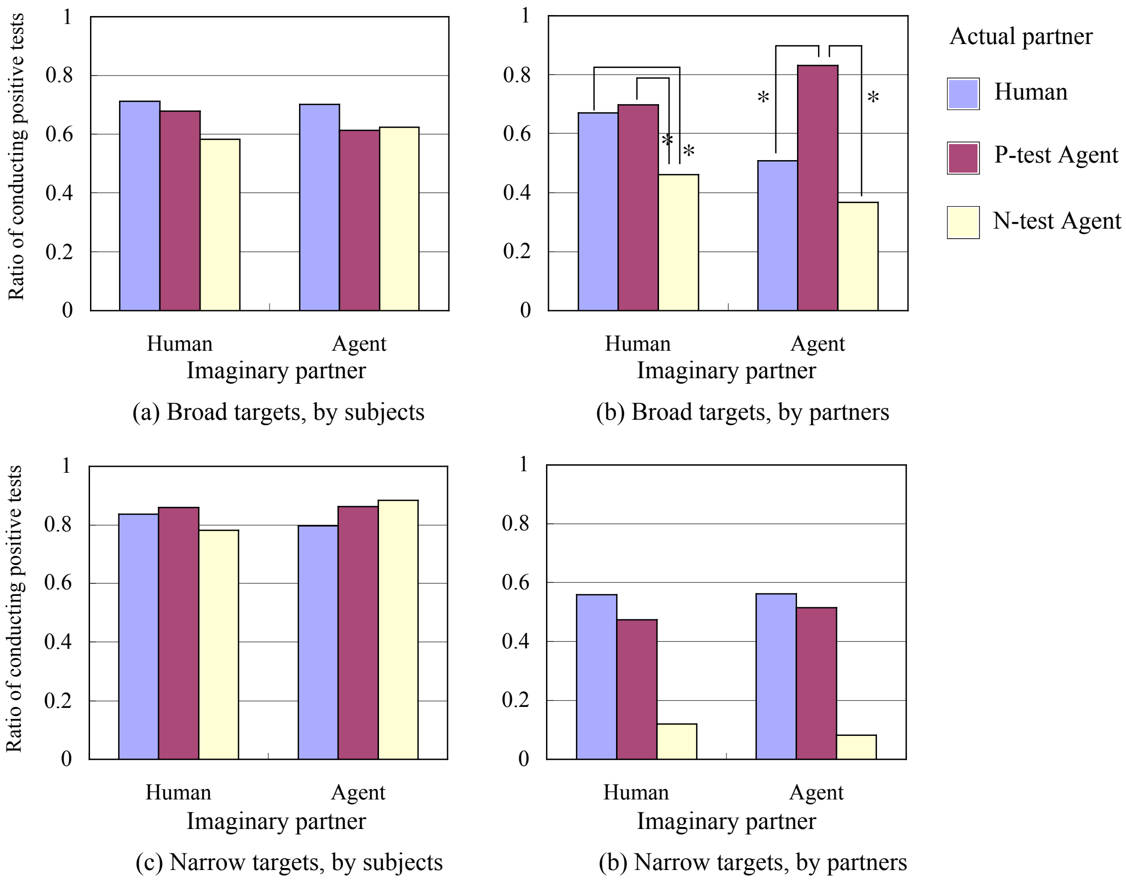

Figure 3 Ratio of conducting positive tests.

This point becomes more interesting when we compare collaboration with the P-test Agent and collaboration with the N-test Agent, where the partner's hypothesis testing strategy was controlled. Figures 3(b) and (d) show that in collaboration with the P-test Agent, the ratio of the partner's instances fulfilling the positive test for the subjects' hypothesis was higher than in collaboration with the N-test Agent. A 2 (*imaginary partners*) x 3 (*actual partners*) ANOVA revealed that the interaction between the two factors was significant in Fig. 3(b) ($F(2, 90) = 4.21$, $p < 0.05$) and the significant difference by a LDS analysis is indicated by a "*" in the figure ($MSe=0.0455$, $p < 0.05$). The same ANOVA reveals that the main effect of an *actual partner* was significant in Fig. 3(d) ($F(2, 90) = 35.87$, $p < 0.01$), and a LDS analysis indicated that the ratios in the Human and P-test Agent conditions were higher than that in the N-test Agent condition ($MSe=0.0537$, $p < 0.05$). Neither the main effect of an *imaginary partner* 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 an *actual partner's* hypothesis testing strategy, this did not influence the subjects' positive test bias. Moreover, this consistency did not depend on the experimenter's instruction as to whether the subjects collaborated with a human subject or with a computer agent (i.e., the change of an *imaginary partner*).

Hypothesis formation

Laughlin & Futoran (1985) indicated that in group activities an individual accepts other group members' hypothesis as his/her own hypothesis while estimating the validity of the others' hypotheses accurately, and this brings about the superiority of group activities to individual activities. Next, we discuss how subjects' reference to a partner's hypothesis in their hypothesis formation is influenced by the change of 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 2 (*imaginary partners*) x 3 (*actual partners*) ANOVA revealed that the main effect of an *actual partner* was significant in finding the broad target ($F(2, 90) = 7.71$, $p < 0.01$), and a LSD test showed that the ratios in the Human and P-test Agent conditions were higher than that in the N-test Agent condition ($MSe=0.0407$, $p < 0.05$). Neither the main effect of an *imaginary partner* 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 an *imaginary partner*: $F(1, 90) = 1.50$, $p > 0.1$; the main effect of an *actual partner*: $F < 1$; the interaction: $F < 1$). This means that in such cases the subjects' tendency to adjust their hypothesis to the partner's hypothesis became stronger. This tendency did not depend on the experimenter's instruction as to whether the subjects collaborate with a human subject or with a computer agent (an *imaginary partner*).

In cases of collaboration with a computer agent, the algorithm in the agent's hypothesis formation was consistent; therefore the subjects were presented with similar hypotheses under the P-test Agent and N-test Agent experimental conditions. However, it is interesting that the subjects tended to adjust their hypothesis to the partner's more remarkably when collaborating only with the P-test Agent. As Fig. 2 showed, the subjects felt that collaboration with the P-test Agent was more familiar than collaboration with the N-test Agent; this tendency may bring about the tendency of strong adjustment to hypotheses given by the positive test agent.

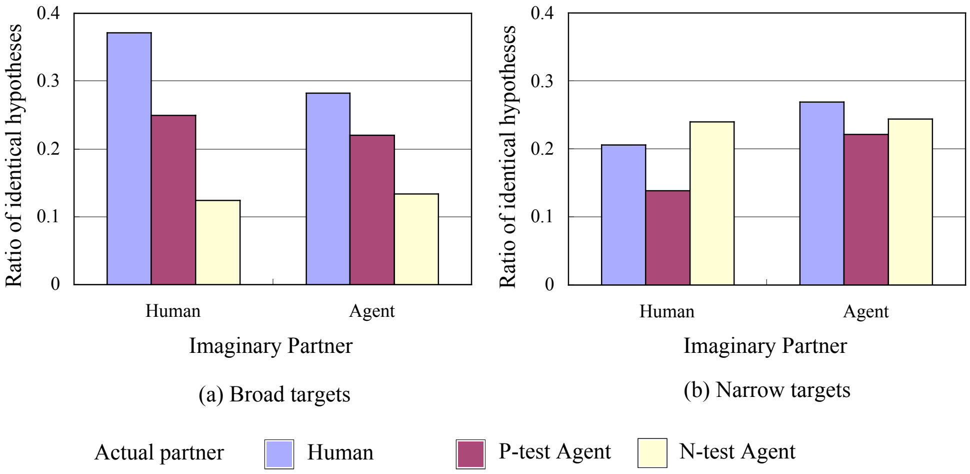

Figure 4 Ratio of adjustment of subjects' hypothesis to the partner's.

CONCLUSIONS

The experimental result shows:
(1) a bias appearing in human behavior, such as the positive test bias in hypothesis testing, was not influenced by the change of an *actual partner*;
(2) the degree of using information given by a partner such as reference to a partner's hypothesis varied considerably with the change of an *actual partner*.

Neither phenomenon above depended on type of *imaginary partner* that was provided by the experimenter.

In finding the narrow target, there was no tendency of subjects adjusting their hypothesis to the P-test Agent's hypothesis (see Fig. 4(b)), whereas there was a tendency found in finding the broad target (see Fig. 4(a)). Why did this difference emerge?

In finding the broad target, the possibility of the agent receiving a Yes as feedback in the experiment was much higher than that 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. Therefore, in finding the narrow target, the P-test Agent faced many *false positives* by receiving a No feedback, which repeatedly rejected the agent's hypothesis. On the other hand, in finding the broad target, the P-test Agent faced many *positive hits* by receiving a Yes feedback, confirming its hypothesis many times (Klayman & Ha, 1987; Miwa, 2004). From the viewpoint of the subjects who observed the agent's activity, this means that in finding the broad target, the P-test Agent seems to propose a reliable hypothesis whereas in finding the narrow target, it usually proposed a dubious one. This difference brought about the result that only in finding the broad target did the subjects tend to adjust their hypothesis to the agent's hypothesis.

REFERENCES

Baylor, A. L. (2000). Beyond butlers: intelligent agents as mentors. Journal of educational computing research, 22, 373-382.

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.

Moreno, R., Mayer, R. E., Sires, H. A., and Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? Cognition and Instruction, 19, 177-213.

Morishima, Y., Nakajima, H., Brave, S., Yamada, R., Maldonado, H., Nass, C., and Kawaji, S. (2004). The role of affect and sociality in the agent-based collaborative learning system. Lecture notes in computer science, 3068, 265-275.

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.

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.