Compound effects of expectations and actual behaviors in human-agent interaction: Experimental investigation using the Ultimatum Game

Yugo Hayashi (y-hayashi@acm.org)
College of Comprehensive Psychology, Ritsumeikan University
2-150 Iwakura-cho, Ibaraki, Osaka, 567-8570, Japan

Ryo Okada (lt0601hf@ed.ritsumei.ac.jp)
College of Letters, Ritsumeikan University
56-1 Kitamachi, Toji-in, Kita-ku, Kyoto, 603-8577, Japan

Abstract

This study investigated how the expectations of others (i.e., top-down processes) and actual perceived behavior (i.e., bottom-up processes) influence negotiations during human-agent interactions. Participants took part in several sessions of the ultimatum game; we investigated the bargaining strategies directed toward the computer agent. To investigate the influence of top-down and bottom-up processes on performance, we designed an experiment wherein (1) participants expected their partners were humans or agents, and (2) agents used different types of algorithmic behavior. Results revealed that irrational decisions, which are characteristic of human-human interactions, emerged when participants believed their opponents were human and when opponent behaviors were ambiguous. Further, we found participants adopted different bargaining strategies according to their expectations and the agent's specific algorithmic behavior. We discuss interplay of the two types of cognitive processing in human-agent interaction.

Keywords: human-agent interaction; top-down/bottom-up processes; social interaction; ultimatum game

Introduction

Studies in human-computer interaction have revealed that how people engage with systems depends on how the agents are perceived (Nass, Moon, Fogg, Reeves, & Dryer, 1995). The human user responds to social cues (Johnson, Veltri, & Hornik, 2008) and to the apparent level of agency of the system (Blascovich et al., 2002). Studies have focused on how users adaptively interact based on their developing representation of the agent, which can be driven by the use of prior knowledge, such as using heuristics (top-down processing), and which can be modified based on the agent's actual behavior (bottom-up processing) (Hayashi & Miwa, 2008). However, it is still unclear how the interdependence of these cognitive processes emerges, and it is not fully understood in which situations such interdependence occurs. To investigate these issues, we conducted a human-agent experimental study that involved negotiation in an ultimatum game.

Two types of cognitive processing in human-agent interaction

Under what circumstances human-like traits such as agency are assigned to computers has been investigated in the fields of human computer interaction and interfaces (Kiesler, Waters, & Sproull, 1996; McEneaney, 2013; Nass et al., 1995; Johnson et al., 2008; Blascovich et al., 2002). Theoretical studies of human computer interaction (e.g., Nass et al.

(1995)) have noted that people unintentionally respond to technology that exhibits social traits as if it were human, as a way to conserve cognitive resources and maximize response efficiency. HCI studies also suggest that how people perceive computers depends on the social cues that are designed into the system. For example, human facial features (Gong, 2008), embodied gestures (Buisine & Martin, 2007), and language use (McLaren, DeLeeuw, & Mayer, 2011) provide for a human-like agent that evokes social responses. However, there is controversy associated with this theory: such automatic responses have been suggested to be aberrant behaviors that result from situational inattention or inappropriate overgeneralization (McEneaney, 2013).

Recent studies in human-agent interaction (HAI) have pointed out the importance of top-down and bottom-up cognitive processing (Miwa & Terai, 2006). Top-down processing is based on the socialized knowledge of others, i.e., interpersonal schemas or stereotypes (Fisk & Taylor, 1991). Such processing is essential for developing representations of others in the initial stage of interaction, and can be used as supplemental information when representations are difficult to develop based on other's behaviors. However, the representation of others may change over time due to their ongoing behavior and the context in which the interaction occurs (Hayashi & Miwa, 2008). Such behavior-based processes are examples of bottom-up processing.

It is important to note that in interpersonal communication between humans, people flexibly use both types of cognitive processing to economically process information when developing representations of others and deciding upon a response. However, few studies have investigated the relationship between the two types of processing in HAI, and it is unclear how such processing plays a role in interactions. Accordingly, in this study we used the Ultimatum Game (UG), a bargaining game that is commonly used in behavioral economics (Guth & Tietz, 1990), to investigate how the combination of expectations and actual behavior influences cognitive processing during decision making.

Influence of top-down and bottom-up processing in an ultimatum game

The ultimatum game is often used to investigate behaviors that are not self-regarding, such as choice inequity and reciprocity (Yamagishi et al., 2009). This game is played by two

players a *proposer* and a *responder*. Typically, one individual actively participates at any given time (i.e., it is a turn-taking game).

First, the proposer receives a sum of money from the experimenter and then makes a proposal concerning how to share the money with the responder. The responder is given two alternatives, namely to either reject or accept the proposal. If the proposal is accepted by the responder, both players receive money according to the proposal, but if the responder rejects the proposal neither receives any money. As such, the self-regarding profit-motivated behavior is to accept any proposal.

Interestingly, respondents tend to reject proposals that are not distributed fairly, even when doing so results in a loss of profit for both players (Guth & Tietz, 1990). In the current study, it is assumed that if the respondent (participant) perceives the proposer (agent) as human, the former may react accordingly, such as by rejecting proposals and abandoning profit as in human-human studies. We controlled the expectations (i.e., top-down processing) of the participants and determined whether expectations of their partner, such as believing the partner is human or non-human, would produce irrational behavior.

H1: When given an unfair proposal, the rejection rate by the respondent will increase when he/she thinks the partner is human compared to a computer agent.

However, as mentioned previously, actual behavior during interactions is used to update the representation of others (i.e., bottom-up processing). To investigate this issue, we used a multi-period version of the ultimatum game (mUG) (Guth, 1995). Studies have revealed that over repeated trials, players learn to expect that the proposer will suggest a fair deal in some future trial; as such, proposal rejections tend to decrease. That is, the number of rejections decreases due to understanding the strategy of the opponent (Slembeck, 1999). Therefore, we hypothesized that if the agents (proposers) showed concessional bargaining behaviors, and participants could perceive such behavior, respondents would perform more rationally by reducing the frequency of rejections.

H2: The rejection rate will decrease when participants understand that the proposer will provide concessional proposals.

Assuming that top-down and bottom-up processing are interdependent, it can be further assumed that the effect of expectations will emerge only when others' behaviors can be explicitly interpreted. To investigate this issue, we produced agents with different algorithmic behaviors, which will be described in more detail in the following section.

Method

Participants and procedure

Seventy-six (male: 30, female: 46, *M* age: 21.38, SD: 1.03) Japanese university students majoring in psychology voluntarily participated in the task; 3 were subsequently excluded

from data analysis because they discovered that their partner was not human.

Participants collected in small groups in a computer room and were instructed how to play the mUG game. They were told that they would play the role of either the proposer or responder; however, all were actually assigned the role of responder and the computer agent played the role of the proposer.
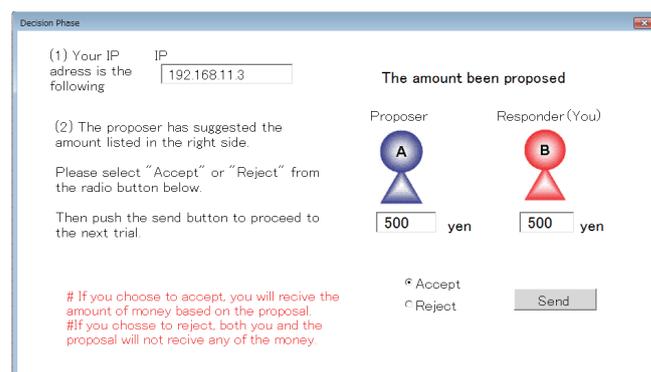

Figure 1: Example screenshot the task.

After the brief introduction to the task, participants were told to start the program, which appeared to connect to a randomly chosen peer in the computer room. They were told that 1,000 Japanese yen (approximately 12 dollars) was provided to the proposer. On the left hand side of the screen, the participant was required to input his or her IP address, which was nominally for connection to the opponent. Below were simple instructions including what he or she would/would not receive based on his or her decision. On the right hand side, the current proposal was shown. Below were decision buttons and a send button to transmit the result to the proposer.

First, a screen appeared that prompted the participant to wait until the proposer finished entering the amount of the proposal. After a short delay, the screen changed to that shown in Figure 1. Then, the participant chose to either *accept* or *reject* the proposal.

A proposal and subsequent decision constituted one trial and a total of 15 trials were conducted in one set of this task; two sets of this task were conducted in total. After completing the task, the participant wrote down a description of how he or she felt about his or her partner.

Experimental conditions

This study examined mUG performance changes due perception of the partner as human or non-human and the partner's actual behavior. We used a 2 (perceived partner: human vs agent) X 4 (actual behavior: random vs adaptive [simple, egocentric, exocentric]) experimental design. The perception of the partner was controlled by telling the participant that the partner was either human or a computer agent. The former was called the *human* condition and the latter the *agent* condition.

In each set of the task, the announcement that the partner was human or computer was announced and the order of such announcements was counterbalanced between the small groups. There were no differences in rejection rates according to the order.

To investigate the effect of agent behavior, we implemented agents that utilized (1) algorithmic behavior or (2) no such algorithmic behavior (random condition). To determine how participants change their interactive strategies based on perceived behavior and ongoing interactions, we implemented three different types of behavior for (1). We examined different algorithms, including those that were likely to be perceived as offering generous or fair proposals. If bottom-up processing predominated in this task, the participant would likely adopt the rational strategy of accepting all proposals from these types of agent.

Behavior of agent

In this section we describe the parameters that defined the agent behaviors. Table 1 shows all possible responses that could be generated by the agent for each trial. In the first trial, the agent always selected response type 4 in all conditions. Then, in the next and subsequent trials, the probability of generating each different response type differed according to the condition.

Table 1: Types of response(proposals) by the agent

response type	amount of proposal (yen)	
	proposer(agent)	responder(participant)
r1	100	900
r2	200	800
r3	300	700
r4	500	500
r5	700	300
r6	800	200
r7	900	100

In the random condition, the agent selected fair/unfair proposals (response type 1-7) randomly, with equal probability. This allowed for the investigation of ambiguous behaviors (i.e., restricting bottom-up processing). In egocentric, exocentric, and adaptive conditions, the agent proposed concessional and generous responses based on the participant's decisions.

In the simple adaptive condition(hereinafter referred to as adaptive condition) the agent repeated the proposal if it was accepted, and otherwise proposed the completely opposite monetary strategy (i.e., fair versus unfair). This was based on the *Pavlof* strategy in social games, wherein the basic rules are "win-stay" and "lose-shift" (Nowak & Sigmund, 1992). In Figure 2, *SAME* denotes repeating the same proposal as in the prior trial.

The egocentric and the exocentric conditions were based on the adaptive condition. In the egocentric condition, the agent responded such that the proposal was clearly biased toward the computer agent(see Figure 3). More specifically, the

$[r1 - r3]$
"accept" $- > \%SAME\%$
"reject" $- > r5 - r7 : 33.33\%$
$[r4]$
"accept" OR *"reject"* $- > r1 - r3, r4 - r7 : 16.66\%$
$[r5 - r7]$
"accept" $- > \%SAME\%$
"reject" $- > r1 - r3 : 33.33\%$

Figure 2: Algorithm schematics of the adaptive condition.

agent reacted economically, such as proposing r3 if the participant kept accepting this proposal. The agent behavior in the egocentric condition is shown below. The agent decided on the next proposal depending on whether the participant *accepted* or *rejected* the previous proposal. For example, in the first trial the agent always proposed r4 (see Table 1). On trial 2, if the participant selected *accept*, then the agent generated the next proposal based on the following probabilities: r1 (10 %), r2 (20 %), r3 (70 %).

$[r1]$
"accept" $- > r1 : 10\%, r2 : 20\%, r3 : 70\%$
"reject" $- > r5 : 10\%, r6 : 20\%, r7 : 70\%$
$[r2]$
"accept" $- > r2 : 30\%, r3 : 70\%$
"reject" $- > r5 : 10\%, r6 : 20\%, r7 : 70\%$
$[r3]$
"accept" $- > $ **r3 : 100**$\%$
"reject" $- > r5 : 10\%, r6 : 20\%, r7 : 70\%$
$[r4]$
"accept" $- > r1 : 10\%, r2 : 20\%, r3 : 70\%$
"reject" $- > r5 : 10\%, r6 : 20\%, r7 : 70\%$
$[r5]$
"accept" $- > r5 : 10\%, r6 : 20\%, r7 : 70\%$
"reject" $- > r1 : 10\%, r2 : 20\%, r3 : 70\%$
$[r6]$
"accept" $- > r6 : 30\%, r7 : 70\%$
"reject" $- > r1 : 10\%, r2 : 20\%, r3 : 70\%$
$[r7]$
"accept" $- > $ **r7 : 100**$\%$
"reject" $- > r1 : 10\%, r2 : 20\%, r3 : 70\%$

Figure 3: Algorithm schematics of the egocentric condition.

In the exocentric condition the agent responded such that it sought less profit than in the egocentric condition(see Figure 4). If the participant kept accepting the proposals, the agent gradually proposed r1 more frequently, and even unfair, agent-biased proposals were most often r5 (i.e., relatively modestly favoring the agent).

```
[r1]
"accept" − > r1 : 100%
"reject" − > r5 : 70%, r6 : 20%, r7 : 10%
[r2]
"accept" − > r1 : 70%, r2 : 30%
"reject" − > r5 : 70%, r6 : 20%, r7 : 10%
[r3]
"accept" − > r1 : 70%, r2 : 20%, 3 : 10%
"reject" − > r5 : 70%, r6 : 20%, r7 : 10%
[r4]
"accept" − > r1 : 70%, r2 : 20%, r3 : 10%
"reject" − > r5 : 70%, r6 : 20%, r7 : 10%
[r5]
"accept" − > r5 : 100%
"reject" − > r1 : 70%, r2 : 20%, r3 : 10%
[r6]
"accept" − > r5 : 70%, r6 : 30%
"reject" − > r1 : 70%, r2 : 20%, r3 : 10%
[r7]
"accept" − > r5 : 70%, r6 : 20%, r7 : 10%
"reject" − > r1 : 70%, r2 : 20%, r3 : 10%
```

Figure 4: Algorithm schematics of the exocentic condition.

Results

Performance of participant: Rejection rate

The participants' percentage rejections are shown in Figure 5. The vertical axis represents the average percentage of proposals rejected during the 15 trials, the horizontal axis shows each behavioral condition, and the different bar shading denotes the different instructions.

A 2 instructions (human or agent) x 4 agent behaviors (random, adaptive, egocentric, or exocentric) mixed factorial ANOVA revealed a significant interaction between the two factors ($F(3, 72) = 4.535$, $p = .0057$). Analysis of simple main effects indicated that in the random condition, proposals by an apparently human opponent were rejected more often than those of a computer opponent ($F(1, 72) = 18.144$, $p = .0001$), whereas there were no differences for the adaptive, egocentric, and exocentric conditions ($F(1, 72) = 0.504$, $p = .4800$; $F(1, 72) = 2.016$, $p = .1600$; $F(1, 72) = 0.165$, $p = .6862$, respectively).

The simple main effect of instruction (human or agent) was also significant for each behavior condition ($F(3, 72) = 9.543$, $p = .0001$; $F(3, 72) = 3.388$, $p = .0198$). Multiple comparisons using Ryan's method for the human instruction and showed that rejections were higher for the random condition than the adaptive, egocentric, and exocentric conditions ($p = .0001$; $p = .0001$; $p = .0076$, respectively). For the agent instruction, the random condition only differed from the egocentric condition ($p = .0052$). Also, when they were instructed that their partners were agents, the egocentric condition was associated with less rejections than the exocentric condition ($p = .0092$).

To summarize, the effect of instruction was significant when the behavior of the agent did not have any intention (i.e., the agent engaged in non-adaptive behavior). This indicates that H1 is supported only when others' behaviors cannot be used to understand their strategy (i.e., bottom-up processing is not possible). In contrast, the effect of the behavior markedly influenced the participants' performance; therefore, H2 is supported. However, participants' performance changed contingent on how they perceived their partner. That is, instruction and behavior interacted.
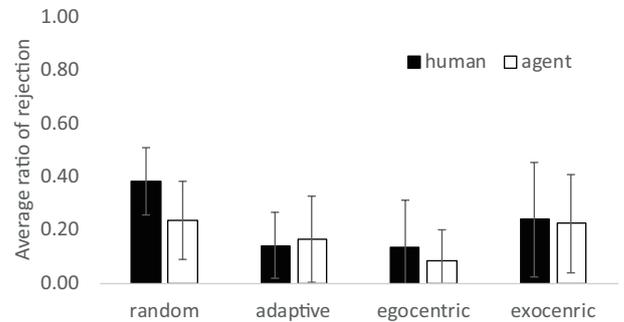

Figure 5: Ratio of rejections.

Behavior of agent: ratio of proposal types

To further understand how the agents adaptively changed their behavior due to the participants' decisions, we examined the actual proposals made by the agents. Figure 6 shows the distribution of proposals for each condition. We then conducted an ANOVA that included the three behavioral conditions that adaptively changed their behavior based on the participants' decisions.

For the human condition, we conducted a 7 x 3 mixed factorial ANOVA with the seven selected responses (r1, r2, r3, r4, r5, r6, or r7) and adaptive conditions (adaptive, egocentric, or exocentric) as independent factors. There was significant interaction between the two factors ($F(12, 324) = 22.147$, $p = .0001$). Since we wanted to investigate which response appeared most frequently within each condition we only conducted simple main effects analysis for each level of condition. Significant main effects were present for all conditions (adaptive: $F(6, 324) = 5.211$, $p = .0001$; egocentric: $F(6, 324) = 45.798$, $p = .0001$; exocentric: $F(6, 324) = 18.403$, $p = .0001$).

Next, multiple comparisons using Ryan's method were conducted for the adaptive condition. Response types r1, r2, and r3 were used more frequently than r3, r4, r5, and r6 ($p = .0001$, for each comparison). For the egocentric condition, response r3 was used more often than all other responses (r1, r2, r4, r5, r6, and r7; $p = .0001$, for each comparison). For the exocentric condition, response r1 was chosen more frequently

than r2, r3, r4, r6, and r7 ($p = .0001$, for each comparison) and response r5 was used more frequently than r2, r3, r4, r6, and r7 ($p = .0001$, for each comparison).

For the agent condition, we conducted the same analysis and found a significant interaction between the two factors ($F(12, 324) = 27.581$, $p = .0001$). Focusing on the same simple main effects, responses differed according to condition ($F(6, 324) = 4.541$, $p = .0001$; $F(6, 324) = 52.996$, $p = .0001$; $F(6, 324) = 22.469$, $p = .0001$). Multiple comparisons revealed exactly the same pairwise differences were significant as in the human condition ($p = .0001$, in each case).

To summarize: (1) in the adaptive condition, r1, r2, and r3 were used most frequently; (2) in the egocentric condition, r3 was most commonly used; and (3) in the exocentric condition, r1 and r5 were the most frequent proposals. This shows that agents responded differently to the participants' decisions and that the agent frequently generated proposals that did not favor itself in the exocentric condition.
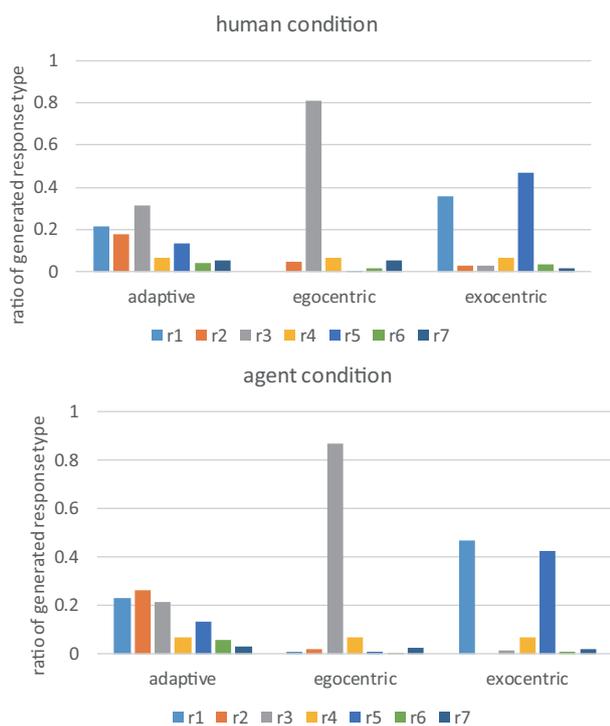

Figure 6: Ratio of generated proposal(top: human condition, bottom:agent condition).

Discussion
Influence of the expectation of the other

The rejection rate data revealed that when the opponent had no strategy (i.e., random condition) the effect of expectations played an important role (i.e., human condition vs. agent condition). This shows the influence of top-down and bottom-up processing and their interdependence, whereby participants used initial expectations to generate a representation of their opponent when the opponent's behavior was not clearly interpretable.

However, why did participants reject the proposer's offer most frequently when the proposer was believed to be human? Past research on economic behaviors using the UG has provided various explanations as to why participants reject proposals, even when doing so is not rational (Guth & Tietz, 1990). Fehr and Schmidt (1999) proposed "inequity aversion theory," which posited that people are sensitive to unfair proposals, regardless of who profits most. People aim to balance inequities by rejecting unequal proposals. Furthermore, Falk, Fehr, and Fischbacher (2003) suggested that following unfair proposals, rejections will rise due to the interpretation of how the proposal was decided upon. I.e., there is an attribution of intentionality or animosity by others. As such, participants may have attributed the same types of intentions to their opponent in this study. However, when they believed their opponent was non-human, such human-specific effects did not occur and rejections decreased.

Influence of the types of adaptive behaviors

The agent's behavior strongly affected rejection rates, whereby participants tended to reject proposals less frequently when the opponent adopted consistent and adaptive strategies, compared to the inconsistent random condition. This tendency was most pronounced when the partner was believed to be human. This indicates that participants decided upon a strategy based on their understanding of the adaptive behavior (i.e., using bottom-up processing), but relied on initial expectations (i.e., using top-down processing) when the opponent's behavior was unpredictable.

Interestingly, participants tended to behave more rationally (i.e., accepting the proposals) when they expected to interact with an agent only when the agent used an egocentric strategy. This indicates that expectations of such egocentric agents may have suggested that the system was non-negotiable toward fairer proposals, and thus the best strategy was to accept their proposals.

Surprisingly, compared to the egocentric condition, participants behaved more irrationally in the exocentric condition by rejecting proposals that were beneficial to them, such as r1. Figure 6 shows that participants oscillated between r1 and r5 as a consequence of their pattern of rejection and acceptance of proposals. However, why did they reject proposal r1? This can be interpreted as rejection to reduce the dissonance (Festinger, 1957) associated with an unfair proposal, regardless of who profits. Further, this could be a result of adopting social norms, such as inequity aversion (Fehr & Schmidt, 1999). Such a socially interactive approach may be the result of perceiving the agent as a social actor (Nass et al., 1995).

These findings cast new light on how decisions in human-agent interaction change based on the compound effects of who an actor believes his or her opponent is, and the actual behavioral strategy observed.

Conclusions

This study investigated the influence of top-down (i.e., expectations of others) and bottom-up processing (i.e., the observation of human-like strategic behavior) on human-agent interaction. This aim was to determine the interdependence of such processing, and to investigate how these processes influence rational decision making in a mUG.

Based on evidence that people reject unfair proposals in human-human interactions, we hypothesized that believing one's partner is human will influence the rejection of other's proposals, if the other's intentions are difficult to interpret (i.e., bottom-up processing cannot be used). By conducting a virtual human-agent experiment, we controlled participants' expectations via agent behavior that followed simple algorithms. The results supported our hypothesis and show that people rely on expectations of the opponent's behavior when the latter's actual behavior is ambiguous. This highlights the interdependent relationship of top-down and bottom-up processing in human-agent interaction.

In addition to the effects of the two types of processing in the mUG, results suggest that people try to avoid inequity; that is, to reject unfair proposals even if they are profitable for themselves. Such a tendency was observed here, even when the participant believed their opponent was a computer agent. This indicates that people treat their counterparts as social actors, even when the goal of the interaction is self-regarding.

In summary, this study supports the interdependent influence of two types of cognitive process, and captures the emergence of irrational decision making in human-agent interaction.

References

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103-124.

Buisine, S., & Martin, C. J. (2007). The effects of speech gesture cooperation in animated agents' behavior in multimedia presentations. *Interacting with Computers*, *19*(4), 484-493.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, *41*(3), 20-26.

Fehr, E., & Schmidt, K., M. (1999). A theory of fairness, competition, and cooperation. *Oxford JournalsSocial Sciences Quarterly Journal of Economics*, *114*(3), 817-868.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fisk, T. S., & Taylor, E. S. (1991). *Social cognition*. McGraw-Hill Education.

Gong, L. (2008). How social is social responses to computers? the function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, *24*,(4), 1494-1509.

Guth, W. (1995). On ultimatum bargaining experiments? a personal review. *Journal of Economic Behavior and Organization*, *27*(3), 329-344.

Guth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*(3), 417-449.

Hayashi, Y., & Miwa, K. (2008). Schema-based and evidence-based communication in human-human and human-agent interaction. In *Proceedings of 6th international conference of cognitive science* (p. 285-288).

Johnson, R. D., Veltri, N. F., & Hornik, S. (2008). Attributions of responsibility toward computing technology: The role of interface social cues and user gender. *International Journal of Human Computer Interaction*, *24*(6), 595-612.

Kiesler, S., Waters, K., & Sproull, L. (1996). A prisonor's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology*, *70*(1), 47-67.

McEneaney, E. J. (2013). Agency effects in human-computer interaction. *International Journal of Human-Computer Interaction*, *12*(29), 798-813.

McLaren, M. B., DeLeeuw, E. K., & Mayer, E. R. (2011). Polite web-based intelligent tutors: Can they improve learning in classrooms? *Computers and Education*, *56*(3), 574-584.

Miwa, K., & Terai, H. (2006). Analysis of human-human and human-computer agent interactions from the viewpoint of design of and attribution to a partner. In *Proceedings of the 28th annual conference of the cognitive science society* (p. 597-602).

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human Computer Studies*, *43*(2), 223-239.

Nowak, A. M., & Sigmund, K. (1992). Tit for tat in heterogenous populations. *Nature*, *355*(6357), 250-253.

Slembeck, T. (1999). Reputations and fairness in bargaining: experimentalevidence from a repeated ultimatum game with fixed opponents. In *University of st.gallen, technical report* (p. 1-21).

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. In *Proceedings of the national academy of sciences of the united states of america pnas* (p. 11520-11523).