DOI: http://dx.doi.org/10.7551/978-0-262-31709-2-ch099

Evolution of Mutual Trust Protocol in Human-based Multi-Agent Simulation

Hirotaka Osawa[1] and Michita Imai[2]

[1]University of Tsukuba
[2]Keio University
osawa@iit.tsukuba.ac.jp

Abstract

Acquisition of the opponent's model and achieving mutual trust with others are notable traits of humankind's intelligence. Achieving mutual trust is a big challenge for artificial intelligences, and it is a key factor in trading. However, how players observe each others' behaviors and how they achieve mutual trust are not fully known. In this study, we researched the growth of a mutual trust protocol in a trading game in a human-based simulation. We designed and implemented web-based multi-player trading game based on the refusable iterative Anti-Max Prisoner's Dilemma game (rAMPD). In the game, each agent's strategy is described by an automaton and periodically modified by human players. We conducted a long-term human-based evolution of mutual trust using this trading game for approximately one month and observed how the agents' automata changed. Analyses of the high-ranking agents' automata and introspective reports by the human players revealed that the mutual trust protocol is achieved by using the initial trade as a signal for mutual recognition.

Introduction

Intentional reading by an agent is an important topic in the field of artificial life. Learning the other's intention is called the Theory of Mind (ToM), and under the social brain hypothesis, it is thought to be a main factor in the evolution of our brains (Premack & Woodruff, 1978) (Byrne & Whiten, 1989). Being able to estimate the intentions of other people and trust them are important factors in trading in the real world and require intelligence. The "power of trust" becomes larger if an agent's reward is maximized or its penalty is minimized through trading; i.e., trading is encouraged when a winning agent gets a large reward and the losing one incurs only a small loss. Fisher and Shapiro used iterative arm wrestling for teaching the importance of trust in trading (Fisher & Shapiro, 2005). They demonstrated that if two players play an iterative arm wrestling game and the winner gets a reward in each match, it is better for both players to fix the game rather than engage in a real fight. They also showed that the key factors in agreeing to fix a game is that each player needs to be intelligent and trust that after if he or she intentionally loses a match, his or her opponent will intentionally lose the next match. They showed that mutual trust sometimes emerges even without words being exchanged between players.

The Iterative Prisoner's Dilemma (IPD) is a typical game in game theory, and it is designed in such a way that the reward is maximized if both players cooperate (Axelrod, 1984). A cooperative strategy in the IPD is achievable without players having to estimate each other's strategy. This kind of game model is appropriate for simulating ecological behaviors of animals that do not relate to ToM (Le & Boyd, 2007). On the other hand, it is insufficient for representing mutual trust in trading situations because mutual trust requires delayed actions. A human player can lose an arm wrestling match and still believe his or her opponent a will lose in the next match.

The current study is on a human-based multi-agent simulation of a refusable iterative Anti-Max Prisoner's Dilemma game (rAMPD). It was conducted to see how mutual trust in trading arises. The Anti-Max Prisoner's Dilemma (AMPD) was first proposed by Angeline (Angeline, 1994). He modified the reward table of IPD so that it could cover the mutual trading behavior of Fisher and Shapiro's iterative arm-wrestling game. We included refusal as the third choice of the agent in AMPD (hence, refusable AMPD, or rAMPD). This extension can simulate real-world trading because each player has the right to ban opponents in free trade. We recruited 74 people to play the rAMPD in a simulation lasting 28 days, and the results of our analyses show how mutual trust arose during this game.

The following sections are organized as follows. Section 2 explains game rule of rAMPD. Section 3 explains how we implemented the system for human-based evolution and conducted experiments and the result of the experiment is shown in section 4. Section 5 analyzes the result and discusses how mutual trust and other strategies are acquired by agents. Section 6 describes how our result contributes to other research field and section 7 describes our method's limitation. Section 8 concludes the paper.

Game Rules

Table 1 is the reward table of the trading game. The standard IPD conditions are shown in equation 1, and the AMPD conditions are shown in equation 2.

Table 1. Reward table of Trading Game

B \ A	Cooperate	Defect
Cooperate	$(A:c, B:c)$	$(A:a, B:b)$
Defect	$(A:b, B:a)$	$(A:d, B:d)$

$$a > c > d > b, \quad a + b < 2c \quad (1)$$
$$a > c > d > b, \quad a + b > 2c \quad (2)$$

We also added 'refusal to trade' as a choice for each agent. If the refusal is selected by an agent, the two agents finish their trade with no chance of retrying. We modified Axelrod's reward table (a = 5, b = 0, c = 3, d = 1) because it is commonly used in game theory simulations. To increase the value of refusal selection, we averaged four constants. We subtracted 3 from c and subtracted 2 from the remaining three. The reward table for rAMPD was thus (a = 3, b = -2, c = 0, d = -1). The average of the four constants was 0, and this satisfied equation 2. The value c (=0) represents an example of Fisher and Shapiro's arm wrestling that both player's cooperative hands do not make sense.

In the rAMPD game, all agents traded with each other iteratively. We also selected the maximum matches in one trade up to 100 times. All agents traded in a round robin fashion. The round robin was repeated several times. The human participants could improve their agent's strategy between each round robin.

Human-based Evolution

Notation of the strategy by automaton

Each participant got his/her own agent and input strategy of the agent through an automaton. We selected automaton-based description of strategy in three reasons. First, automaton-based strategy is understandable to participants especially who are not familiar about programming. Second, the automaton is easy to analyze because of its simple notation. Third, the automaton has enough describable for complex strategy.

Each participant input their agent's strategy by using a finite state automaton. Each state in the automaton had numbers representing cooperation and defection of the agent. Even states represented cooperation, odd states represented defect, and 0 represented a refusal. The transition arrows between states were described with triplets numbers. The first number represents the present state, the second number represents the opponent's hand (0 means cooperate and 1 means defect), and the third state represented the next state (even, odd, or 0 state). Each participant described their strategy using the start state number and several triplets. For example, {{2}, {2,0,2}, {2,1,2}} means a strategy that is anytime cooperative. {{1}, {1,0,1}, {1,1,0}} means coward exploiter. If it is once attacked, it refuses trade. {{2}, {2,0,2}, {2,1,1}, {1,0,2}, {1,1,1}} shows the strategy of tit-for-tat which is famous in IPD. We found that the finite state automaton made it easy for players to understand each others' behaviors, and that it is enough descriptive to maintain mutual trust.

Implementation of the game: how to motivate participants

For motivating participants, we designed the simulation as an online game. All games were implemented in AJAX style, and participants input their strategies using a web form shown in Fig. 1. Each participant could download his/her agent's trade history from the website at any time (shown in Fig. 1 top). The results of a trade were calculated on the server side and feedback to participants both ranking page and interactive result viewer. Each participant could replay their previous result in viewer mode (shown in Fig. 1 bottom).

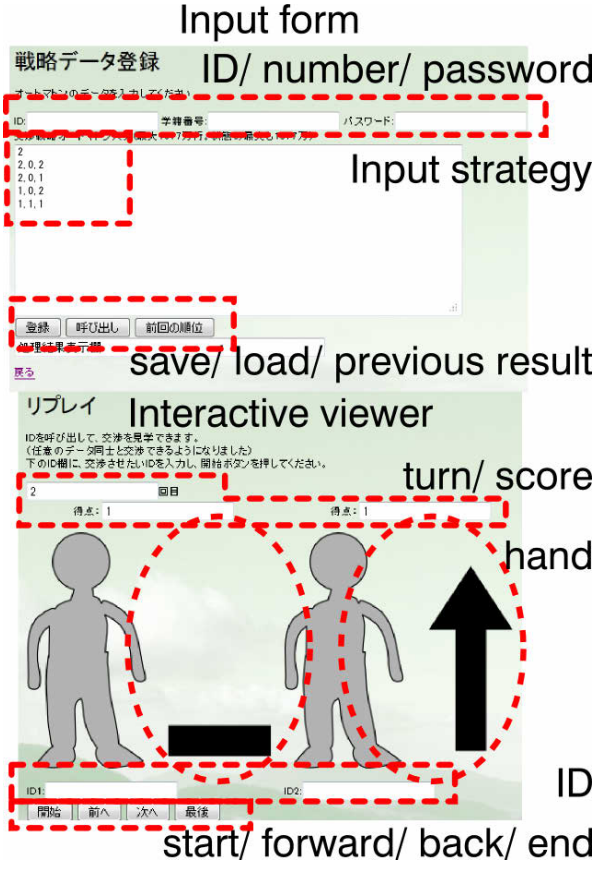

Figure 1. Implemented game screen. Top figure shows the input form of the automation code of the text. Bottom figure shows the viewer mode. Both figures are captured from web browser.

We also wrote a cover story for the game to nurture the imagination of the participants and motivate their play. In the cover story, the participants were residents of an island, and they traded fish in a poisoned pond. Each fish became edible if it had been dipped in a different pond. Each agent could choose between three selections, to wait at home (C), to go to

an opponent's home and take fish (D), or to lock the door (refusal). If both agents wait at home, there was no reward and no penalty (0, 0). If both went the other's home, the doors were locked and both became tired (-1, -1). If one went to the opponent's home and the opponent was waiting there, he or she could eat the fish and the opponent lost it (+3 ,-2). If one of the players refused, the communication and trading stopped. This story is a bit artificial. However, every participant understands how the rule works.

Participants

We conducted an experiment in a class learning about automata, and the participants were students from that class. In total, 74 participants played the game. The experimental period ran from 2012/5/29 12:30 to 2012/6/26 8:30. Trading was conducted four times a day during breaks between classes (8:30, 12:30, 16:30, and 20:30), and the ranking table was updated during this period. The chances for updates totaled 112 times. Because updates were done during breaks, each participant had enough time to input strategies and confirm the update's result before and after trading.

All participants were given scores for their class work according to their ranking at the end of the simulation. We divided up the participants who took more than 1 point into 16 groups (these agents were survivors which ate fish and were not hungry). Each group member got from 20 points to 5 points in order to his/her agent's score. The rules were described to the participants before the game started.

It was important for us to confirm that there were no ethical problems in conducting this experiment as a part of the automaton class because it was designed as both an experiment and as a means for students to learning the basic behaviors of automata. The experiment also included an evaluation of the students.

Results

There were 1109 updates of the automata. The average number of update accesses per trade was 9.9. The average number of updates per player was 15 times. 68 agents achieved more than 1 point at the time of the last update, and their programmers were given bonus points in the class. According to their acquired score, we named each agent in order of highest score A01 to lowest score A74, categorized the 68 participants whose agents exceeded 1 into groups 20 (G20) to 5 (G5), and put 6 participants with less than 0 points in group 0 (G0). The average length of the agents' automata was 33.7. The average length of automata with more than 1 point was 36.5. Figure 2 shows the average length of automaton and average points in each group.

We categorized each player's state using the following rules. If both players each got more than 40 points and less than 60 points, we considered that both players trusted each other and categorized them into the mutual trust (MT) group. If one of the players got more than 50 points and the other got less than -50 points, we considered that one of the players exploited (EX) the other player and that other player was exploited (EXd). If both players got less than -50 points, we considered that both players could not trust each other and mutual destruction (MD) occurred. If both players got less than 10 points and trading was stopped by one of the players, the trading was banned (BA). If both players got less than 10 points and trading continued until the end of the simulation, trading resulted in stagflation (ST). All categories are shown in Fig. 3. ST only happened in the lower-rank group G0. There are trends on Fig. 3 that higher-ranked agents achieve more mutual-trust than lower-ranked agents (note that BA in high-ranked agents are still required to prevent lower-ranked agents' attacks).

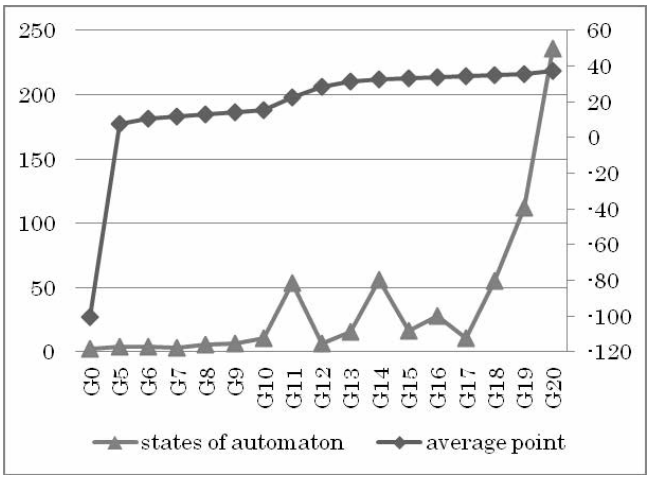

Figure 2. Average states of automaton and average points for each group. The right axis shows the number of automaton states, and the left axis shows the average number of points. The bottom axis shows the 17 groups.

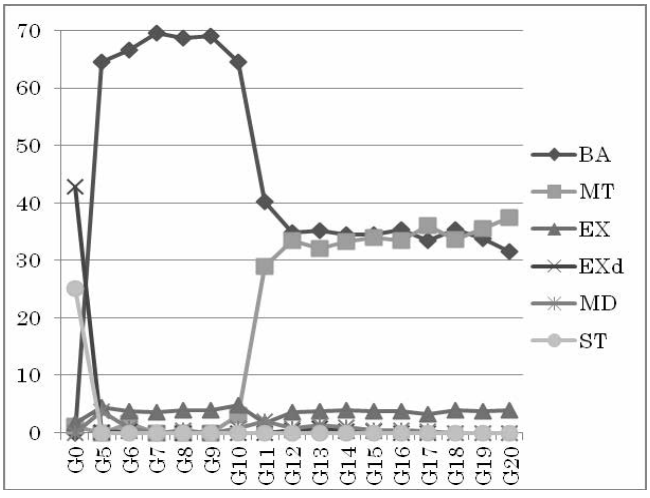

Figure 3. Average number of categories in each group. The bottom axis shows the 17 groups.

Discussion

The General strategy of the high-ranked agent

Figure 4 shows a strategy of the high-ranked agent as a meta automaton. All high-ranked agents (in the G20, G19, and G18 groups) had four phases of their strategy. First, the agent repeated cooperate or defect in a determined order. This determined order was different in each agent. If the opponent selected a different hand, the agent transited to the mutual trust phase. In the mutual trust phase, the agent tried to take a complementary hand. If the opponent selected C continuously, the agent transited to the exploiting phase and tried to exploit the opponent. On the other hand, if the agent detected D continuously, the agent transited to the refusal phase and finished the trade

The detailed transition rules depended on the agent. Note that in this game, two identical automata do not succeed because they cannot change to different hands. This restriction discourages users from cheating and accelerates the evolution of the identification process.

From the participants' reports, we confirmed that participants gradually came to understand the several dilemmas in this game. For example, if the identification process is too strict (using lots of confirmation before mutual trust), the opponent may regard that it is impossible to cooperate and simply refuse the trade. This loses the chance of possible cooperation. However, if the identification is too loose, the opponent may think that the agent is too foolish to cooperate and start to exploit it. This also loses the chance to cooperate and reduces rewards.

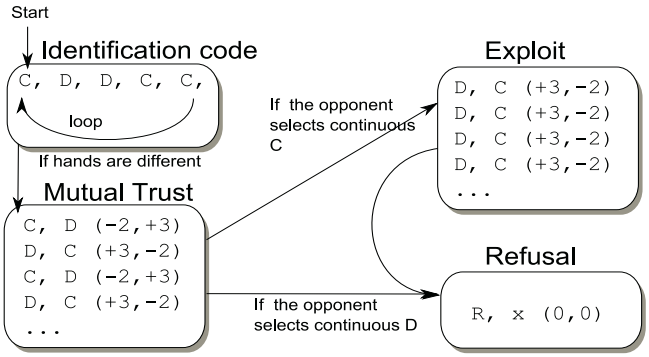

Figure 4. Example of strategy in high-ranked agent (the identification code is extracted from A02)

We studied the results of these three methods by conducting a statistical analysis, manual analysis (where the author input hands manually, traded with each agent, and observed the behaviors), and collecting reports from the participants. The next subsection discusses the analyses conducted during each phase.

Identification Code Phase

We confirmed that at least 30 agents (A01-A30) had an identification code phase of manual analysis and the participants' reports. Each agent in 30 agents has a different set of hands on the start process (CDDCC, DCDC, etc.). If each agent's hand is different, they start to go mutual trust phase.

The length of the identification code loop was less than five pairs (For example, A02 had 5 loops and A04 had 3 loops). Theoretically, a 2^5 bit unique code is required to identify 32 agents. This result corresponds to the fact that the agents that had identification phases numbered less than 30.

57 participants selected D as the first hand of the agent and 17 participants selected C. Most participants selected D because the agent had a chance to get a higher score than the opponent. On the other hand, the participants who selected C reported that selecting C in the first hand had an advantage because it was easier to start up a mutual trust situation with it than D. One of the participants reported that his strategy followed an old proverb "win by losing."

Mutual trust phase

Theoretically, mutual trust arises in a longer loop, like CCDD, CCCDDD, etc. However, all agents used short mutual trust loops (CDCDCD...). 42 agents (A01-A41 and A61) had mutual trust phases. As shown in Fig. 3, high-rank agents had more mutual trust (MT) and fewer refusals (BA). This result suggests that lower-rank agents lost the chance through their own or their opponent's (BA) refusal, whereas high-ranked agents could use the chance to make a mutual trust loop (MT). A larger set of states in an automaton weakly suggests that mutual trust requires each automaton to have more complex states. This result suggests the validity of the social brain hypothesis wherein evolution of intelligence (approximated by the number of states) is accelerated through identification in society (Byrne & Whiten, 1989).

Exploiting Phase

The exploiting phase was observed in almost all agents (A01-A71). There were 4 exploited agents (A71-74, all agents were in G0). Continuous Cs was a trigger for transiting to the exploiting phase. We confirmed that all agents who had an exploiting phase transited to the exploiting phase after 2 to 4 continuous Cs. As shown in Fig. 3, almost all agents attacked during the transit from the exploiting phase to the refusal phase.

Keeping the exploiting phase has a risk in that it may imprudently trigger the opponent's refusal phase. However, the exploiting phase was preserved until the end of the game because there are big rewards for exploiting phase. There were three participants who input only C regardless of their opponent's hand and did not change strategies (A72-A74). A71 had a simple strategy that selected a CCDCCD... loop unrelated for the opponents' hands. These weak agents kept being exploited.

Refusal Phase

The refusal phase was inevitable because if the opponent selected continuous D regardless of its hand, the agent prevented a negative score just by selecting refusal (C,D and D,D are both negative). 41 agents had a refusal phase (A01-A41).

All of these agents reacted to more than 3 continuous Ds. On the other hand, several agents allowed 1 or 2 Ds. These behaviors kept the opponent cooperative and not "be anger" (for triggering opponent's refusal attitude).

Difference between rAMPD and AMPD

Each agent acquired more complex strategy in rAMPD compared with a strategy in AMPD game proposed by Angeline (Angeline, 1994). There are three factors for generating different results from previous research.

The first factor is difference of simulation. Angeline uses computer-based simulation for evolving each agent's strategy. On the contrary, we used human-based simulation. The latter condition expands the possibility of each agent's strategy. The second factor is the description of the strategy. Angeline uses set of four hands (CCCC-DDDD) for describing strategy of each agent. On the other hand, we used automaton for describing strategy. The automaton makes it possible to use more complex strategy. The third factor is the refusal hand of each agent. Refusal phase creates "point of no return" in each trade and it makes communication complex. In Angeline's AMPD game, each agent has no refusal selection and a trade continues to determined cycle. If an opponent plays continue Ds (which means two or more continued D hand), most appropriate strategy is replying with continuous D. If the opponent stops continuous D, the agent just needs to stop continuous D. In our situation, a most appropriate hand for the opponent's continuous D is just refusal. However, if continuous D are produced mistakenly by the opponent, there is still a chance for creating mutual trust in each other. We hypothesized that the refusal phase is the critical factor for evolving complex communication between agents. As a future work, we will confirm the hypothesis by using computer-based simulation.

Contributions

We tried a human-based multi-agent simulation instead of a computer-based one. The human-guided approach is used in several fields, from artificial life, cloud sourcing, and human interfaces (Kosorukoff, n.d.)(Paolacci et al., 2010)(Osawa & Imai, 2012). Our results suggest that this approach also works if the motivations of the human players are carefully designed.

Our findings revealed two important factors related to game theory and multi-agent simulation. The first is in regard to the emergence of mutual trust in trading itself. In game theory, the possibility of mutual trading can be analyzed in the Cheap Talk Game that divides a trading game into an initialization phase and a main trading phase (Wärneryd, 1991). Our results suggested that identification of others and mutual cooperation can be achieved even without “cheap talk" by using the reward itself. This finding may lead to multi-agent simulations becoming simpler as far as their requirements go. The second factor is the importance of being able to refuse during free trade. Previous studies mainly focused on the locality of the agent as a way of avoiding agents they were not confident about and this leads to agents forming clusters (Suzuki & Arita, 2001). This approach is good for ecological simulations. However, general free trade is not dependent on the distance to the others, but rather on the mutual intention of trading. Our results suggest that agents come to believe each other and reject agents they are not confident about not by using additional information (like cheap talk and location) but rather through behaviors.

In light of the above discussion, we think that our human-based multi-agent simulation of the rAMPD game reflected the essence of real-world free trading and gave us good insights about how identification of others and mutual trust arises in humans.

Last, we want to note that human-based multi-agent simulation quickly proceeds analysis for the game space because we can collect agent's process of evolution by participants' introspections. We want to emphasize that most participants are motivated by this gamification method (students involved in our "homework" make good scores in class). We believe that motivated participants are very good research factors for estimating the possibility of game space especially in earlier stages of study.

Limitations

Human-based simulations are dependent on capricious humans. To handle human resources properly, we need to design the experimental setup carefully.

Although there were 112 updates in this task, almost all of them happened during the first week and final week. To maintain motivation during the whole game period, we may need to back-reward the participants (for example, by scoring on a weekly basis).

Three participants did not update their agents, and this sabotage influenced the other participants. Several participants complained in the report that the authors did not evict these three agents. We think that these variable motivations also reflected a real simulation. However, this result also shows the importance of a good motivation design in human-based multi-agent simulation. This underscores the need to carefully design the agent's goal - each participant's motivation - in a human-based multi-agent simulation.

Knowing the number of trades may increase the unwanted factors . The top scoring groups (G20 and G19) had more than 100 states in their automata. The analyses of the automata and the reports from the participants showed that a large number of states were prepared for the 100th match. Defect or refusal is the optimal strategy even if mutual trust arises because the 100th match does not have a succeeding match. There were also three agents that prepared the 100th match in G11, G14 and G16, as the spikes in Fig. 1 show. The reward (<5 points) for defect or refusal in the 100th match was relatively small

compared with the points from MT (around 50 points) and EX (around 300 points). The main ranking seemed dependent on the amount of mutual trust and the 100th match did not influence mutual trust. The increasing trend in MT in going from G10 to G20 (Fig. 2) supports this idea. These unwanted evolutions can be avoided if the number of matches is indefinite in each update.

Human-based simulations sometimes encounter ethical problems. For example, we could not regulate communications between participants, unlike in the case of a computer based multi-agent simulation. In this experiment, the participants were rivals and there was no real motivation to cooperate. Moreover, cloning was meaningless in this task. These two facts restricted communication between participants. The participant reports also suggested that there was no cooperation between the participants. However, it is hard to monitor the sorts of strategy that could have been generated through discussions with other participants. This problem may be avoided if the game is conducted online anonymously and all behaviors are monitored. Anyway, the experimenter must be careful about regulating human behaviors. It is important to ensure that the experiment is profitable for the participants themselves.

Conclusion

We designed and implemented a web-based multi-player trading game based on the refusable iterative Anti-Max Prisoner's Dilemma game (rAMPD). In this game, each agent's strategy is described by an automaton and is periodically modified by human players. We conducted a one-month human-based multi-agent simulation using this trading game lasting approximately one month and observed how the agents' automata changed. Analyses of high-ranking agents' automata and introspective reports from the human players revealed that a mutual trust protocol arises using the initial trade as a signal for mutual recognition.

Acknowledgements

Our work was supported by the diligent students at Keio University. This work was supported by the JST PRESTO program.

References

Angeline, P. J. (1994). An Alternate Interpretation of the Iterated Prisoner's Dilemma and the Evolution of Non-Mutual Cooperation. *Proceedings of 4th artificial life conference* (pp. 353–358).

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Byrne, R. W., & Whiten, A. (1989). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, USA.

Crawford, V. P., & Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, *50*(6), 1431 – 1451.

Fisher, R., & Shapiro, D. (2005). *Beyond Reason: Using Emotions as You Negotiate* (p. 256). Viking Adult.

Kosorukoff, A. (n.d.). Human based genetic algorithm. *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)* (Vol. 5, pp. 3464–3469). IEEE. doi:10.1109/ICSMC.2001.972056

Le, S., & Boyd, R. (2007). Evolutionary dynamics of the continuous iterated prisoner's dilemma. *Journal of theoretical biology*, *245*(2), 258–67. doi:10.1016/j.jtbi.2006.09.016

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*(6398), 826–829. doi:10.1038/359826a0

Osawa, H., & Imai, M. (2012). Possessed Robot : How to Find Original Nonverbal Communication Style in Human-Robot Interaction. *International Conference on Agents and Artificial Intelligence (ICAART)* (pp. 632–641). INSTICC.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419. doi:10.2139/ssrn.1626226

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526. doi:10.1017/S0140525X00076512

Suzuki, R., & Arita, T. (2001). Evolutionary Analysis on Spatial Locality in the N-person Iterated Prisoner's Dilemma. *Proceedings of inaugural workshop of artificial life*, 105 – 114.

Wärneryd, K. (1991). Evolutionary stability in unanimity games with cheap talk. *Economics Letters*, *36*(4), 375–378. doi:10.1016/0165-1765(91)90201-U