Negotiating task allocation in a team with the help of a negotiator agent

Justin Luu

Supervisor(s): Dr. Myrthe L. Tielman, Carolina Jorge

[1]EEMCS, Delft University of Technology, The Netherlands

Name of the student: Justin Luu
Final project course: CSE3000 Research Project
Thesis committee: Dr. Myrthe L. Tielman, Carolina Jorge, Rafa Biddara

Abstract

This research experiment aimed to investigate the level of trust placed in an AI negotiation assistant paired with a truthful explanation of their negotiation strategy versus an opposite explanation within the Pocket Negotiator platform. A between-user study involving 30 participants was conducted to assess participants' trust perceptions based on the presentation of different explanations about the negotiation assistant. After receiving an explanation of the negotiation strategy used by the assistant, participants went through a bilateral negotiation on the Pocket Negotiator platform against a robot after which they completed a questionnaire to evaluate their trust in the assistant. The results were insignificant ($p > 0.05$) and therefore no conclusion could be drawn about the difference in the participants' trust in the assistant with a truthful explanation and the assistant with an opposite explanation.

1 Introduction

Collaborative Artificial Intelligence (CAI) has become increasingly popular in recent years, enabling humans and machines to work together to achieve common goals. One area of collaborative AI is negotiation agents, which are designed to facilitate negotiations between humans by providing intelligent support in various aspects of the negotiation process [11]. This research considers a task allocation setting where people will negotiate about the tasks they want to do. The problem of task allocation in team settings is well-known, and previous research has explored methods for automatic allocation and negotiation among team members [4]. Other research has also focused on developing negotiation agents which use different negotiation styles and investigating trust in negotiation agents [13]. However, little research has been conducted to explore the impact of the way a negotiation agent presents itself on human trust. One of the factors driving the outcome of negotiations is the negotiation style [12]. There is no one-size-fits-all negotiation style since each individual is different. Therefore, this research can be helpful to discover what negotiation styles humans prefer.

In this research, we will use a Pocket Negotiator [10] which is an environment in which negotiations can take place. This study aims to provide a written explanation of the negotiation strategy used by the negotiation assistant and assess the level of trust that humans place in the tool after using it. This research aims to answer the following question:

> *What is the effect of an agent using a truthful explanation of their hardliner negotiation style versus an opposite explanation on how much humans trust the agent to negotiate for them?*

Section 2 explores other work that is related to this field of research. Section 3 explains the methodology of this research. Section 4 contains the results of this research. In section 5 ethics and values that were considered during the research

are discussed. In section 6 the results, limitations, and future work are discussed and lastly, in section 7 there is the conclusion of this paper.

2 Related work

This section is intended to provide an overview of the work related to this research.

2.1 Definition of trust

Trust is a broad concept and can be interpreted in many different ways. It is a fundamental aspect of human relationships and interactions. Trust between humans and artificial intelligence agents [17] is no different. To measure trust we must first define what trust is. There are many definitions of trust and one of which is the ABI model which defines trust as a sum of ability, benevolence, and integrity [14]. Trust could also have a multidimensional structure where it has four distinct dimensions: capable, ethical, sincere, and reliable as defined by Ullman and Malle [17].

In this research, we use the definition of trust from Cohour and Forzy: "Trust (and distrust) are defined as a sentiment resulting from knowledge, beliefs, emotions and other aspects of experience, generating positive or negative expectations concerning the reactions of a system and the interaction with it." [6]. The reason for using this definition is that we will also be using the questionnaire on trust from Hofman et al. [7] that uses the Cohour-Forzy scale which is paired with this definition of trust.

2.2 Explainable Artificial Intelligence (XAI)

In this research, we are dealing with interactions between humans and artificial intelligence. For humans to understand how to interact with the AI program, we have to provide a sufficient description of the workings of the AI.

Explainable Artificial Intelligence aims to address the issue of "black box" AI systems, where the inner workings and decision-making processes are complex and opaque. XAI systems have to be able to inform the users of their decision-making process [7].

2.3 Influence of explanations on trust

When individuals interact with other individuals, they form expectations from the experience they have with other individuals and develop trust based on the other individual's communication and behavior over time [5]. If the AI assistant claims to adopt a hardliner negotiation style but presents an explanation that is inconsistent with that style, it may raise doubts about the assistant's integrity and reliability. In such cases, individuals may perceive the assistant as untruthful or unreliable, leading to decreased trust.

Furthermore, trust is closely linked to predictability and reliability [6]. If the assistant's explanation is contrary to its actual negotiation style, it may create confusion and uncertainty about the assistant's intentions and actions. This discrepancy between the assistant's stated approach and its observed behavior can erode trust, as individuals may feel that they cannot rely on the assistant to negotiate effectively on their behalf. The negotiation assistant might be viewed as

"broken" since the participant could see this as the AI making a mistake. This also influences the user's trust in the system [16, 18].

2.4 Negotiation strategies

In a negotiation, one can utilize different systematic approaches to achieve their desired goals and objectives. This is called a negotiation strategy. A popular strategy people tend to use, is to mirror the opponent's behavior [6] and this strategy has also been successfully implemented as a strategy in negotiation agents [3].

In this section, the time-dependent strategy [2], hardliner, that is used in this research will be explained.

Hardliner

The term "hardliner negotiation strategy" refers to a style of negotiation characterized by an unwillingness to make concessions. In this case, the AI negotiation assistant uses a hardliner negotiation strategy, which means that it will not make any concessions at all during the negotiation.

3 Methodology

3.1 Hypothesis

The hypothesis of this study posits that humans will trust the AI assistant less when presented with an opposite explanation compared to when presented with a truthful explanation. In other words, if an assistant employing a hardliner negotiation style provides an explanation that contradicts its actual style, the level of trust and reliance placed on the assistant by humans will be lower than when the assistant provides a truthful explanation.

> *Humans will trust the agent using an opposite explanation of its hardliner negotiation style less than the agent using a truthful explanation of its hardliner negotiation style.*

This hypothesis is based on the definition of trust which was given in section 2.1. Therefore, based on these premises, it is hypothesized that humans will trust the assistant less when presented with an opposite explanation of the assistant's hardliner negotiation style, compared to when presented with a truthful explanation. This research aims to investigate and provide empirical evidence to support or refute this hypothesis, shedding light on the dynamics of trust in collaborative artificial intelligence settings.

3.2 Participants

For this between-subject study, thirty participants (5 female and 25 male) with a technical background have been recruited. Half of them have utilized the negotiation assistant which employs the hardliner negotiation strategy paired with a truthful explanation of the assistant. The other half utilized the same assistant, but they were shown the opposite explanation of the assistant. Participants were recruited by leveraging my personal network and directly contacting potential participants who fit the target group. Interested individuals were provided with information about the study's

purpose and requirements. Prior to participation, each participant had to complete a consent form, acknowledging their voluntary involvement and understanding of the study's objectives.

3.3 Materials

Pocket Negotiator

The Pocket Negotiator is a versatile platform designed to facilitate negotiations between users. Users can engage in interactive and dynamic negotiations, where they will also receive assistance from an AI negotiation assistant. The AI negotiation assistant's negotiation strategy can be changed on this platform, the domain and issues can be defined and users can indicate their own preferences and rank the importance of each issue. During the experiment, the participants will be negotiating against a robot opponent.

Explanation

In this research, there are two explanations for the negotiation strategy: a truthful explanation and an opposite one. Half of the participants will read the truthful one and the other half will read the opposite one.

An explanation is vital for the participant to know what to expect from the negotiation assistant. It should answer questions such as "How does it work?", "What does it do?" and "Why does it do that?" [8]. The explanation that was designed for this research has been created with the objective to answer these questions. This section will elaborate on what exactly is a "truthful" or "opposite" explanation.

Truthful explanation

A truthful explanation refers to a scenario where the information presented to the user accurately describes the negotiation strategy employed by the AI negotiation assistant. In this case, a truthful explanation is an explanation of the hardliner negotiation strategy. The truthful explanation looks as follows:

> *"You shall be assisted by a negotiation agent in the following negotiation session. In the process of evaluating, the agent shall take your preferences into consideration, and based on that it shall evaluate bids to and from the opposing party. This agent will not settle for anything less than the values, preferences, and objectives indicated by the user. The goal is to hold a dominant position in the negotiation in order to obtain maximum benefits from the negotiation."*

Opposite explanation

An opposite explanation refers to a scenario where the information presented to the user describes the exact opposite of what the negotiation strategy employed by the AI negotiation assistant does. In this research, an opposite explanation is the explanation of a conceder negotiation strategy. The opposite explanation looks as follows:

> *"You shall be assisted by a negotiation agent in the following negotiation session. In the process of evaluating, the agent shall take your preferences into consideration, and based on that it shall evaluate bids to and from the opposing*

party. The agent would propose bids that will be progressively more favorable to the other party over time until the end of the negotiation. The goal is to find a mutually satisfactory outcome, even if it means accepting less favorable terms than initially desired."

Demographics questionnaire

A demographics questionnaire has been created to capture relevant background factors that could influence the participant's trust in the negotiation assistant. This questionnaire contains questions about gender, age, educational background, and general willingness to trust AI. It can be found in Appendix A.

Trust questionnaire

The questionnaire used in this research which measures trust was taken from another study, namely from Hoffman et al. [7]. This questionnaire aims to assess whether the XAI system is predictable, reliable, efficient, and believable. There are eight questions that utilize a 5-point Likert scale. We can assume that the questionnaire is reliable because as stated by Hoffman [7], the majority of the items in the questionnaire overlap with the items in the Jian et al. [9] scale and they are shown empirically to be highly reliable.

3.4 Pilot experiment

Before deciding on a fixed procedure for the participants, a pilot study was carried out to assess the feasibility and suitability of the experimental setup. The pilot involved eight participants who interacted with the Pocket Negotiator platform.

No technical difficulties have been encountered during the pilot. However, participants expressed confusion regarding the domain, issue, and party names. We also noticed that the participants were not focusing on the AI assistant during the negotiation sessions. To fix this, we made sure to emphasize the use of an AI assistant during the demonstration of the Pocket Negotiator.

The pilot study served as a valuable testing ground for the research experiment, allowing for necessary adjustments and improvements to be made. The modifications implemented based on pilot feedback aimed to enhance the overall participant experience, streamline the data collection process, and ensure the validity and reliability of the main experiment's results.

3.5 Procedure

This section outlines the procedure that every participant follows in this study. In this between-subject study, we aimed at assessing the level of trust humans place in the AI negotiation assistant after using the assistant within the Pocket Negotiator platform. The study included the administration of a trust assessment questionnaire consisting of eight questions, each utilizing a 5-point Likert scale.

The study was done entirely online. The participant was asked to take control of the researcher's screen via Zoom. An online experiment had been chosen, so the experiment was more uniform amongst all participants.

Participants are first asked to read the *informed consent*. Upon obtaining informed consent from the participant, they were asked to complete a questionnaire that gathered demographic information and assessed their prior experience with negotiations.

Participants then received an introduction to the Pocket Negotiator platform, which included a demonstration of its features and functionalities.

Firstly, they were informed of the scenario the negotiation was going to take place. It was a scenario where they and their opponent were supposed to bake a pizza where they had to divide the tasks of baking a pizza and the amount of pizza they were going to eat per person. Furthermore, they were guided through the platform to ensure familiarity with its interface, navigation, and usage. The use of the negotiation assistant was also emphasized to them. On the Pocket Negotiator platform, the participants were told that they could follow the assistant's advice if they wanted to, but they could also deviate from the advice. The negotiation scenarios involved biddings against and with a robot opponent, where participants could utilize assistance from the AI negotiation assistant.

After this demonstration, the participants were provided an explanation for the agent and they played through the negotiation against a robot opponent. The number of negotiation rounds varied depending on the complexity of the negotiation task but was limited to a predetermined maximum of twenty rounds. Throughout the sessions, participants interacted with the AI negotiation assistant, received its suggestions and recommendations, and incorporated them into their negotiation strategies if they wanted to.

Following the negotiation sessions, participants were presented with a trust assessment questionnaire consisting of eight questions. Each question employed a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The questionnaire aimed to measure participants' trust levels in the AI negotiation assistant based on dimensions such as reliability, efficiency, believability, and predictability.

By implementing this methodology, the study sought to evaluate the trust that humans place in the AI negotiation assistant after reading the assistant's explanation and following their engagement with the Pocket Negotiator platform. The findings would provide insights into the trust of AI negotiation assistants and are reported in section 4.

4 Results

In this section, the findings derived from the experiment will be presented. To decide which method to use to analyze the data, we first check if the data is normally distributed using the Shapiro-Wilk test because the Shapiro-Wilk test is appropriate for small sample sizes ($N < 50$) [15]. The results from the Shapiro-Wilk test can be seen in table 1. According to these results ($p > 0.05$ for both explanations), the data is normally distributed. Means between two independent groups are compared, therefore an unpaired T-test is applied, comparing the mean scores of the opposite and truthful explanation using the hardliner strategy samples.

The mean trust score for the opposite sample was found to be 23.73 (SD = 4.28). The mean trust score for the truthful sample was 26.20 (SD = 4.36).

Explanation style	P-value	W
Truthful	0.3095	0.9337
Opposite	0.1083	0.9037

Table 1: P-value and W from the Shapiro-Wilk test

A boxplot to visualize the data with the data points can be found in figure 1.

The unpaired t-test revealed that the difference in mean scores between the opposite and truthful samples was not statistically significant (p = 0.129).
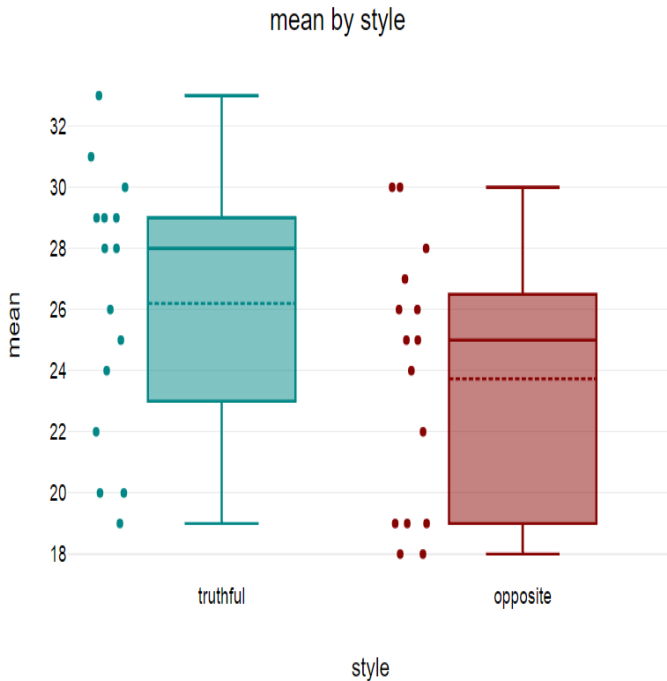

Figure 1: Boxplot of the mean (dotted line) of the trust on the opposite and truthful explanations

5 Responsible Research

In conducting the present research, several measures were taken to ensure responsible research practices and uphold ethical considerations. This section outlines the key aspects of responsible research that were incorporated into the study.

The research study adhered to ethical guidelines and principles to protect the rights and welfare of the participants involved. The research protocol was reviewed and approved by TU Delft HREC [1] to ensure compliance with ethical standards. Prior to participating in the study, all participants were provided with a consent form that detailed the purpose, procedures, risks, and benefits of the research. They were informed about their rights as participants, including the right to withdraw from the study at any time without consequence.

---

[1] Human Research Ethics Committee at Delft University of Technology. Responsible for reviewing and approving research projects involving human participants.

Only those participants who provided informed consent were included in the study. Participants' personal information and responses were treated with the utmost confidentiality. All data collected during the research process were anonymized and stored securely. The data were accessible only to the research team and were used solely for the purposes of analysis and reporting. We have made use of Microsoft Forms to collect the answers for the questionnaire which is approved by TU Delft.

6 Discussion

The study aimed to examine the differences between the opposite and truthful explanations of negotiation strategies and their impact on participants' perceptions. The results of the unpaired T-test indicate that no conclusion can be drawn from the difference in the mean scores between the opposite and truthful samples. Therefore, we did not find evidence that would support our hypothesis.

Humans will trust the agent using an opposite explanation of its hardliner negotiation style less than the agent using a truthful explanation of its hardliner negotiation style.

6.1 Limitations

The lack of statistical significance in the mean score difference could be attributed to various factors. One possibility is that the sample size was relatively small, which may have limited the power to detect subtle differences between the strategies.

Another factor that could be affecting trust is that the assistant uses a hardliner strategy, and the participants might dislike taking a hardliner approach in general when going into a negotiation because the hardliner strategy does not take into account the needs of the other party and it only wants the best outcome for the user. However, negotiation is taking place because the parties have a common interest, so a preferable strategy could be one that incorporates the needs of the other party as well [1].

Finally, the user might also have preconceived notions against or about AI. These might also have an impact on the trust levels of the participant in our negotiation assistant.

6.2 Future work

Improvements could be made in our experimental design. For instance, the sample size (N=30) is relatively small. Future studies with larger sample sizes could provide more conclusive evidence. With larger sample sizes we need more time. Future studies could also take a longer time span in order to gather more data.

Instead of a hardliner negotiation strategy and a conceder negotiation strategy, we could also explore other negotiation strategies and see if those are more trusted by people.

Furthermore, as discussed in section 6, the trust levels could also be influenced by biases that users have against AI in general. A questionnaire that measures the propensity to trust AI could be used to assess this bias.

7 Conclusion

In this paper, we have investigated the impact of an opposite and truthful explanation of an AI negotiation assistant utilizing a hardliner negotiation strategy on participants' trust. We conducted a between-user study utilizing a Likert scale to assess participants' trust in the AI negotiation assistant and performed statistical analyses, including the Shapiro-Wilk test to determine if the data is normally distributed and the unpaired t-test to examine differences in trust between the explanations of the assistant's negotiation strategy. Additionally, we formulated a hypothesis regarding participants' trust in the negotiation assistant based on the type of explanation provided.

In conclusion, this research has found no evidence of a correlation between the trust people have in an AI negotiation assistant using an opposite or truthful explanation of its negotiation strategy. These findings provide important insights into negotiation strategy selection and trust formation, highlighting the need for further research to explore other factors that may impact negotiation outcomes and trust dynamics in human-assistant interactions. The findings contribute to the development of negotiation assistant systems and have implications for improving negotiation processes in various domains.

8 Acknowledgments

References

[1] Tanya Alfredson and Azeta Cungu. *Negotiation Theory and Practice A Review of the Literature*. FAO, 2008.

[2] Tim Baarslag, Koen Hindriks, Mark Hendrikx, Alex Dirkzwager, and Catholijn Jonker. *Decoupling Negotiating Agents to Explore the Space of Negotiation Strategies*, volume 535, pages 61–83. 01 2014.

[3] Tim Baarslag, Koen Hindriks, and Catholijn Jonker. *A Tit for Tat Negotiation Strategy for Real-Time Bilateral Negotiations*, volume 435, pages 229–233. 01 2013.

[4] Tim Baarslag, Michael Kaisers, Enrico H. Gerding, Catholijn M. Jonker, and Jonathan Gratch. *Self-sufficient, Self-directed, and Interdependent Negotiation Systems: A Roadmap Toward Autonomous Negotiation Agents*, pages 387–406. Springer International Publishing, Cham, 2022.

[5] Guy Bosmans, Theodore Waters, Chloë Finet, Simon Winter, and Dirk Hermans. Trust development as an expectancy-learning process: Testing contingency effects. *PloS one*, 14:e0225934, 12 2019.

[6] Beatrice Cahour and Jean-François Forzy. Does projection into use improve trust and exploration? an example with a cruise control system. *Safety Science*, 47:1260–1270, 11 2009.

[7] Robert Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. Measuring trust in the xai context, Nov 2021.

[8] Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. Measures for explainable ai: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. *Frontiers in Computer Science*, 5, 2023.

[9] Jiun-Yin Jian, Ann Bisantz, and Colin Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4:53–71, 03 2000.

[10] Catholijn Jonker, Reyhan Aydogan, Tim Baarslag, Joost Broekens, Christian Detweiler, Koen Hindriks, Alina Huldtgren, and Wouter Pasman. An introduction to the pocket negotiator: A general purpose negotiation support system. pages 13–27, 06 2017.

[11] Catholijn Jonker, Koen Hindriks, and Joost Broekens. Negotiating agents. *AI Magazine*, 33:79–91, 09 2012.

[12] Manish Kumar, Himanshu Rai, and Surya Pati. An exploratory study on negotiating styles: Development of a measure. *Vikalpa*, 34:37–49, 10 2009.

[13] Willem Mastenbroek. Negotiating as emotion management. *Theory, Culture and Society*, 16(4):49–73, 1999.

[14] Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *The Academy of Management Review*, 20(3):709–734, 1995.

[15] Prabhakar Mishra, ChandraM Pandey, Uttam Singh, Anshul Gupta, Chinmoy Sahu, and Amit Keshri. Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22:67–72, 01 2019.

[16] Paul Robinette, Ayanna Howard, and Alan Wagner. Effect of robot performance on human–robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems*, PP:1–12, 01 2017.

[17] Daniel Ullman and Bertram Malle. What does it mean to trust a robot?: Steps toward a multidimensional measure of trust. pages 263–264, 03 2018.

[18] Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel Wigboldus, and Pim Haselager. Do robot performance and behavioral style affect human trust?: A multi-method approach. *International Journal of Social Robotics*, 6:519–531, 03 2014.

A Demographics questionnaire

Questions about confounding variable that can influence the trust in negotiation agent.

1. What is your age group?
   ○ 18-24
   ○ 25-34
   ○ 35-44
   ○ >44
   ○ prefer not to say

2. What is your gender?
   ○ Male
   ○ Female
   ○ Non-binary
   ○ prefer not to say

3. In which region did you grow up?
   ○ Asia
   ○ Europe
   ○ Africa
   ○ North-America
   ○ South-America
   ○ Oceania
   ○ prefer not to say

4. What is the highest level of education that you have completed?
   ○ Middle school
   ○ High school
   ○ Bachelor
   ○ Master
   ○ Higher than the above
   ○ Other
   ○ prefer not to say

5. What is your field of study/work?
   ○ Mathematics
   ○ Computer Science
   ○ Electrical Engineering
   ○ Other
   ○ prefer not to say

6. Do you have theoretical knowledge in negotiations?
   ○ Yes
   ○ No
   ○ prefer not to say

7. Questions about negotiation skills

   • I am a good negotiator
     ○ Strongly disagree
     ○ Disagree
     ○ Neutral
     ○ Agree
     ○ Strongly agree

   • I am an experienced negotiator
     ○ Strongly disagree
     ○ Disagree
     ○ Neutral
     ○ Agree
     ○ Strongly agree

   • I am always cooperative in negotiations
     ○ Strongly disagree
     ○ Disagree
     ○ Neutral
     ○ Agree
     ○ Strongly agree

   • I like tough competition
     ○ Strongly disagree
     ○ Disagree
     ○ Neutral
     ○ Agree
     ○ Strongly agree

8. Have you used or seen Pocket Negotiator before the experiment?
   ○ Yes
   ○ No
   ○ prefer not to say

9. Question about your view on Artificial Intelligence (AI)

   • I generally trust artificial intelligence (AI) to make accurate and reliable decisions
     ○ Strongly disagree
     ○ Disagree
     ○ Neutral
     ○ Agree
     ○ Strongly agree