# Leveraging Social Interactions in Human-Agent Decision-making

Doctoral Consortium

JiHyun Jeong
Cornell Univerisity
Ithaca, NY
jihyun@infosci.cornell.edu

## ABSTRACT

Through social interactions, humans and machines can express their intents, acknowledge each other's, and coordinate with one another to arrive at a joint decision. These interactions can also help achieve goals beyond just task performance. They can help build trust between interactants, which is crucial for effective collaborations. My work aims to i) develop frameworks for social interaction in human-agent joint decision-making and ii) implement artificial agents that improve joint decisions while considering the social and interpersonal implications of their actions. In this extended abstract, I describe past and current work and propose future directions for my dissertation.

## KEYWORDS

Human-agent interaction; Social interaction; Decision-making

**ACM Reference Format:**
JiHyun Jeong. 2021. Leveraging Social Interactions in Human-Agent Decision-making: Doctoral Consortium. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021,* IFAAMAS, 2 pages.

## 1 INTRODUCTION

Decision support systems and algorithm-in-the-loop [6] decisions are appealing to many in various domains. However, an algorithmic output may not be enough for synergistic decision-making with humans partners [5]. These one-off suggestions offer limited interaction opportunities for people to deliberate on decisions with the machine. Hence, we explore leveraging back-and-forth social interactions between humans and agents to make joint decisions.

Increased interactivity can help humans make better decisions with artificial systems. For instance, Elmalech et al. [3]'s work showed that providing incorrect answers that matched human intuition at first resulted in higher receptivity of correct answers, later on, improving average performance over time.

Equipping artificial agents with social roles and capabilities is well-motived in prior work. Artificial agents can serve the social purpose of providing emotional support [9]. Social capabilities offer an opportunity to recover from failures and misunderstandings [7, 13]. Social interactions can also signal a sense of benevolence, one of the core pillars of trustworthiness Mayer et al. [12]. In Bickmore and Cassell [1]'s paper, their embodied conversational agent used small talk as a politeness strategy to build trust.

*Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.*


Trust between partners is integral to effective collaboration. Much like trust between humans, human's trust in machines are based on the machine's ability and the alignment of their intents, motivations, and principles Lee and See [11], Mayer et al. [12]. The latter dimensions become more salient as people may not fully understand the capabilities of increasingly complex systems Lee and See [11].

Against this background, we frame joint decision-making between humans and agents as a constant negotiation of goals and intents while adhering to social norms to maintain relationships. The human and agent both communicate their sets of goals. If there are misalignments in priorities of goals, partners can resolve them and reach an agreement on a shared decision. This process of resolution requires negotiations built on social interactions. Communicating, negotiating, and building consensus are essential parts of the joint decision-making process.

## 2 PRIOR AND ONGOING WORK

Our previous study [10] illustrates how failing to negotiate goals, roles, and strategy, as well as to socially interact are detrimental to collaboration. Often, participants lost trust in the robot that failed to express its priorities and negotiate. They were subsequently more reluctant to accept the robot's recommendations. For example, we observed disagreements over trade-offs when the human and the robot prioritized different goals. Participants also projected maladjusted intent behind the robot's actions, believing that it dismissed them. One noted that although they thought that the robot was making better decisions, they would not want to work with the robot again because they believed it was ignoring their thoughts.

In more recent work [8], we implement an agent that tries to improve the quality of the joint decision while also mitigating frustrations when conflicts arise in negotiations. To demonstrate, we develop a computational framework that models the back-and-forth interactions between the human and the agent. While both the human and agent can suggest, reject, or interchange any options from shortlist, the goal is to reach a consensus on what they think is the optimal decision.

The agent incorporates a type of social ritual in its actions called face-work[4]. Disagreements and harsh criticisms may cause another to suffer a loss of face. To prevent such face-threatening acts [2], the artificial agent may try to compromise with the human's preferences instead of arguing its own. This behavior can happen even at the expense of task performance. The intuition is that an artificial agent might prefer a suboptimal move if the

optimal one has the possibility of ensuing negative emotions that can break a feeling of trust.

Based on this intuition, we present an artificial agent that accounts for face-work in our interaction framework: choosing moves that maximize decision quality unless the same action is a direct face-threat. The artificial agent thus has two goals. One is to improve joint performance. Another, to maintain a good relationship with the human by considering the interpersonal implications of its actions. In this particular scenario, we devise a rule-based method for shifting between the two goals. With our on-going work, we ran experiments with humans to evaluate the framework and the agent behavior.

## 3 FUTURE PROPOSED WORK

Moving forward, I hope to continue to work on improving joint decision-making experiences with artificial agents. Specifically, I propose three future directions for my dissertation.

First, I propose to design intuitive and efficient behaviors for artificial agents to communicate their intentions in negotiations with humans. Clear and expressive capabilities like gestures or explanations can benefit negotiations in decision-making. In particular, clear communication is vital for agents within consequential decision domains such as collaborative search and rescue (SAR) teams or human interactions with autonomous vehicles.

Second, I propose to improve the agent's abilities to select actions that achieve performance and social goals. Instead of the initial rule-based method, agents could use machine learning methods to determine their best course of action. Additionally, agents could make additional inferences or predictions about the human to inform themselves. For instance, the agent might predict humans' receptivity towards its suggestion based on the inferred underlying order of priorities. They could also weigh the effectiveness of social actions to determine the best policy for agent interventions.

Third, much like performance and social goals, I intend to incorporate ethical consideration in choosing agent actions. Ethical principles can guide the design of agent behaviors. There could be specifications and rules for essential principles that agents to adhere to. If rules include competing ethical concerns, the appropriateness of each rule should depend on the domain or the situation. For example, in some cases, it might be irresponsible of the agent to not offer an optimal suggestion when it has one. Agents might not want to manipulate humans who might have to be accountable for their own decisions. In others, good intentions or outcomes (e.g. saving a collaborator's face, improving quality over time) may potentially be worth the deceptive social maneuver. However, if machines are transparent about their social intentions, they could be perceived as overly calculating. Also, there could be a difference in the level of persuasiveness or manipulation that is appropriate depending on how amenable or stubborn the person is.

## 4 CONCLUSION

My work formalizes a framework for human-agent joint decision-making to coordinate preference and priorities while considering performance, interpersonal and social outcomes. The framework will help design social interactions that communicate the agent's intent, incorporate human preferences, and adjust the agent's next set of behaviors. Through iterative modeling, designing, prototyping, and testing of agent algorithms and interventions, I hope to improve its social capabilities and, in turn, the social experience of human-agent joint decision-making.

## ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. W911NF2010004. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA).

## REFERENCES

[1] Timothy Bickmore and Justine Cassell. 2000. how about this weather?" social dialogue with embodied conversational agents. In *Proc. AAAI Fall Symposium on Socially Intelligent Agents*.
[2] Penelope Brown and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*. Vol. 4. Cambridge university press.
[3] Avshalom Elmalech, David Sarne, Avi Rosenfeld, and Eden Shalom Erez. 2015. When Suboptimal Rules.. In *AAAI*. Citeseer, 1313–1319.
[4] Erving Goffman. 1955. On face-work: An analysis of ritual elements in social interaction. *Psychiatry* 18, 3 (1955), 213–231.
[5] Ben Green and Yiling Chen. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, 1–24. https://doi.org/10.1145/3359152
[6] Ben Green and Yiling Chen. 2020. Algorithm-in-the-Loop Decision Making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 13663–13664.
[7] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G Pipe, and Kerstin Eder. 2016. Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 493–500.
[8] JiHyun Jeong and Guy Hoffman. 2020. Face-work for Human-Agent Joint Decision-Making. In *Proceedings of the AAAI 2020 Fall Symposium Series on Trust and Explainability in Artificial Intelligence for Human-Robot Interaction*.
[9] Sooyeon Jeong, Cynthia Breazeal, Deirdre Logan, and Peter Weinstock. 2018. Huggable: the impact of embodiment on promoting socio-emotional interactions for young pediatric inpatients. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[10] Matthew V Law, JiHyun Jeong, Amritansh Kwatra, Malte F Jung, and Guy Hoffman. 2019. Negotiating the Creative Space in Human-Robot Collaborative Design. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 645–657.
[11] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.
[12] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.
[13] Sarah Strohkorb Sebo, Priyanka Krishnamurthi, and Brian Scassellati. 2019. "I Don't Believe You": Investigating the Effects of Robot Trust Violation and Repair. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 57–65.