Socially Intelligent Mission Assistant: Robot Social Norms

Anthony Sanders-Oseguera, Gregory Soukhamneut, Kevin Khamphoumanivong, Noralisa Steiner, Donald Szlachetka, Dr. Boyoung Kim, and Dr. Elizabeth Phillips
United States Air Force Academy

***Abstract* - As technology continues to advance, the United States Military must consider how a robot will be best integrated and accepted as a team member within a unit. The current project examines how the actions of a robot can contribute to its overall acceptance as a team member.**

WHAT

The future vision of robots and other autonomous systems is to function as integrated team members with humans (e.g., Air Force's Autonomous Wingman Program). In fulfilling this goal, it is important for these systems to follow the teams' social norms. Norms provide crucial evidence for the kind of trustworthiness that people will expect from the advanced artificial teammates. In human groups, norms can coordinate behavior (e.g., forming lines while waiting); they increase predictability of others' behavior (because norm-conforming behaviors are known in advance); most importantly, they constrain individuals' self-interest in favor of the group's interest (hence cultivating cooperation); and they foster group cohesion (because efficient, predictable, and cooperative group behavior creates feelings of jointness and belonging). Social Norms can help inform Airmen about the nature and abilities of the robot, which will help them better understand the systems they are working with and be more willing to place trust into the system.

The goal of this work is to investigate how best to implement social and moral norms and ethical frameworks into autonomous systems so they can contribute to team success.

MOTIVATION AND BACKGROUND

Prior researchers have studied autonomous systems that make decisions based on explained reasons and justifications [1]. They have also stated that ethical justification for decision making is imperative for gaining human trust. For instance, systems that can explain their decision making in reference to the social norms of the team are likely to be trusted.

Ramsey discussed how Confucian ethics discusses fundamental commitments to society which can offer an ethical approach of becoming a good team member in society [2]. This role-based ethics emphasizes the importance of adhering to roles that provide decisive reasons for doing most of what they do in their everyday lives [3]. There are two important ideals argued for in role-based ethics: 1) an agent might perform an act of voluntary assumption, such as a promise to perform their role-based duties, which raises expectations in others that they will do the same and 2) people will adhere to role-based ethical decision making because it helps them realize the benefits associated with occupying that role (e.g., Servicemen and women, religious groups, introverts vs. extroverts).

Rule-based ethics, on the other hand, stresses the importance of adhering to rules of behavior because such rules are inherently right or wrong. This may be helpful for robots and other autonomous systems as the rule-based robots may signal that they care more about completing the mission and focused on the law of the mission (rather not the spirit of the mission).

We posit that it is better to have role-based robots versus rule-based robots with ethical implementation to be an effective team member. This is because role-based ethics is primarily focused on being a better team member than deontological ethics.

Before we test this main hypothesis, we first examined how people would respond when a robot engages in morally questionable behaviors. We expected that the findings of this study would shed light on people's overall perception of a robot's capacity to engage in ethical decision-making processes (Study 1). In this paper, we report our preliminary findings from Study 1. In Study 2, we plan to evaluate if humans can trust autonomous robots that have similar social and moral norms as humans.

HOW

Study 1 was a 2 (Agent: Human vs. Robot) X 2 (Cheating: Cheat vs. No Cheat) between-subjects design. Seventy-five participants were randomly assigned to interact with either a NAO robot or a human confederate as an opponent in a game of Rock, Paper, Scissors. The robot was programmed to throw a scripted version of the 30 trials of Rock, Paper, Scissors. After each trial, the robot verbally stating the outcome of the run (e.g., “Aw, you win,” “Yes, I win,” “We have tied this round”). During the study, the participants were video recorded to collect data on their verbal and behavioral responses while playing Rock, Paper, Scissors. Finally, participants were asked to report their perceptions of the robot or the human confederate’s warmth, competence and discomfort experienced through the interaction (The Robotic Social Attributes Scale) [5].

PRELIMINARY RESULTS

Two-way between-subjects ANOVAs were conducted on participants’ warmth, competence, and discomfort ratings, respectively, to examine the effects of agent and cheating on their perception of the robot and the human agents.

Warmth. The analysis of warmth showed that the main effect of agent was statistically significant, $F(1,71) = 166.70$, $p < 0.001$, $\eta^2=0.70$. Ratings of warmth were higher when participants interacted with the human agent ($M = 4.69$, $SD = 1.15$) than the robot agent ($M = 1.44$, $SD = 0.98$).

Competence. The main effect of agent was significant in competence ratings, $F(1,71) = 43.43$, $p < 0.001$, $\eta^2 = 0.35$. Participants judged the agent as more competent when they interacted with the human agent ($M = 5.03$, $SD = 0.89$) than the robot agent ($M = 3.51$, $SD = 1.18$). We also found a significant interaction effect of agent and cheating, $F(1,71) = 6.41$, $p = 0.01$, $\eta^2 = 0.05$. For the human agent, participants’ judgments of competence did not differ in the cheat and the no-cheat conditions. However, for the robot agent, participants judged the robot in the cheat condition ($M = 2.99$, $SD = 1.20$). as less competitive than the robot in the no-cheat condition ($M = 4.05$, $SD = 0.90$).

Discomfort. We found the main effect of the agent to be statistically significant, $F(1,71) = 11.96$, $p < 0.001$, $\eta^2 = 0.14$. Participants interacted with the robot agent ($M = 1.30$, $SD = 1.04$) reported more discomfort than those interacted with the human agent ($M = 0.59$, $SD = 0.65$).

In Study 2, we will further examine whether autonomous robots can serve as an effective team member depending on whether they adhere to role-based versus rule-based moral values.

ACKNOWLEDGEMENT

This work is supported by the National Science Foundation, Cyber Human Systems: Small Collaborative Research: Role-Based Norm Violation Response in Human-Robot Teams, #80NSSC20K0070.

REFERENCES

[1]. Alaieri, F. & Vellino, A. (2016). Ethical Decision Making in Robots: Autonomy, Trust and Responsibility. *Social Robotics Lecture Notes in Computer Science*, 159–168.

[2] Ramsey, J. (2016). Confucian role ethics: a critical survey. *Philosophy Compass, 11* (5), 235-245.

[3] Sciaraffa, S. (2011), Identification, Meaning, and the normativity of social roles. *European Journal of Philosophy, 19* (1), 107-128.

[4] Ramsey, J. (2016). Confucian role ethics and relational autonomy in the Mengzi. *Philosophy East and West, 66* (3), 903-922.

[5] Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017, March). The robotic social attributes scale (rosas): Development and validation. *In Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction* (pp. 254-262). ACM.