ANTHROPOMORPHISM AND TRUST IN HUMAN-AUTONOMY TEAM COMMUNICATION DYNAMICS

Myke C. Cohen, Mustafa Demir, Erin K. Chiou, Nancy J. Cooke
Arizona State University

Project Overview. Communication is a key ingredient of team cognition (Cooke, Gorman, Myers, & Duran, 2013), and is both a factor and a manifestation of trust in human-autonomy teams (HATs; Hou, Ho, & Dunwoody, 2021). Anthropomorphism, or the attribution of humanlike qualities to inanimate objects, is a distinct factor in human trust in automation (Lee & See, 2004). Both anthropomorphism and trust are typically measured through self-report scales; in this exploratory study, we propose a behavioral measure of humans' perceived anthropomorphism of autonomous teammates in the form of *verbal anthropomorphisms*. Examples include the use of gendered and second-person pronouns in reference to autonomy, or the imputation of human emotions and states in communicating with it. We examine the relationship between self-reported and verbal anthropomorphism in simulated remotely-piloted aircraft systems (RPASs) reconnaissance missions under degraded conditions, and compare how they relate to self-reported trust in an autonomous agent.

Design. Three different roles were involved in this study's task: (1) *navigators* responsible for the dynamic flight plan and providing waypoint-related information to the pilot; (2) *pilots* tasked with monitoring and adjusting the altitude, airspeed, effective radius, fuel, gears, and flaps, as well as negotiating altitude and airspeed with the photographer to enable proper conditions for a clear photograph of the target, and; (3) *photographers* in charge of taking clear photos of the target by monitoring and adjusting the camera and providing feedback to the team. A Wizard of Oz (WoZ) paradigm was used, in which the two participants per team were informed that the third member—the pilot—was a "synthetic" agent, when it was a trained confederate mimicking a synthetic agent from a separate room. This "synthetic" teammate used restricted vocabulary to mimic computer language capabilities and facilitate the story that the pilot was a synthetic agent.

The primary study manipulation is the application of three system failures in each mission: (1) *automation failures*, or role level display failures for specific targets, (2) *autonomy failures*, or abnormal behavior of the autonomous agent for specific targets (e.g., providing wrong information to other team members, or misaction), and (3) *malicious cyber-attacks*, or the hijacking of the RPAS, leading to the agent providing false, detrimental information to the team. Each failure was applied to pre-selected target waypoints according to a set schedule, with the malicious cyber-attack appearing only as the last failure of the last mission. Teams had limited time to address each failure, positively related to its difficulty.

Method. A total of 44 participants from a Southwestern US university, split into 22 teams, completed the experiment. They performed ten 40-minute missions distributed across two sessions with a one or two-week interval in between. Each mission had between 12 to 20 targets. For this study, we consider (1) *self-reported measures*, which were Likert-scale questions for trust and anthropomorphism, and; (2) *verbal anthropomorphism*, from chat messages containing anthropomorphizing content, coded in real-time by two experimenters. These included the use of gendered pronouns (i.e., he, she, they), attributing human-like states to the Pilot (e.g., "What do you feel?"), and the use of polite requests directed to the Pilot (e.g., "Please", "Sorry"). Mixed MANOVA was used to compare survey responses across sessions and verbal anthropomorphism levels across missions.

Results and discussion. Trust levels were found to significantly differ for the photographer role across the two sessions, while no significant differences were found in other self-reported measures across role and session, and also for verbal anthropomorphism levels across missions. However, results of pairwise LSD analyses also showed a decline in raw verbal anthropomorphism counts over time, which may indicate a loss of humanlike (and a growth of more tool-like) perceptions of the synthetic teammate as more failures occurred. These findings suggest the plausibility of anthropomorphic content in human communications as a real-time measure of perceived anthropomorphism. Once further established, the relationship between verbal anthropomorphism and trust can be the basis for more robust AI models capable of initiating humanlike trust repair mechanisms for future HAT designs.

Acknowledgements. This work is supported by ONR Award N000141712382 (Program Managers: Marc Steinberg, Micah Clark). We thank Steven M. Shope of Sandia Research Corporation who integrated the synthetic agent and testbed.

REFERENCES

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science*, *37*(2), 255–285. https://doi.org/10.1111/cogs.12009

Hou, M., Ho, G., & Dunwoody, D. (2021). IMPACTS: A trust model for human-autonomy teaming. *Human-Intelligent Systems Integration*. https://doi.org/10.1007/s42454-020-00023-x

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392