



AFRL-AFOSR-VA-TR-2024-0211

**Organizational Implications of
Autonomy-Mediated Interaction**

**Gratch, Jonathan
UNIVERSITY OF SOUTHERN CALIFORNIA
3720 S FLOWER ST FL 3
LOS ANGELES, CA, 90089
USA**

**04/25/2024
Final Technical Report**

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20240425	2. REPORT TYPE Final	3. DATES COVERED	
		START DATE 20180802	END DATE 20230801
4. TITLE AND SUBTITLE Organizational Implications of Autonomy-Mediated Interaction			
5a. CONTRACT NUMBER	5b. GRANT NUMBER FA9550-18-1-0182	5c. PROGRAM ELEMENT NUMBER 61102F	
5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER	
6. AUTHOR(S) Jonathan Gratch			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF SOUTHERN CALIFORNIA 3720 S FLOWER ST FL 3 LOS ANGELES, CA 90089 USA			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR RTA2	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-VA-TR-2024-0211
12. DISTRIBUTION/AVAILABILITY STATEMENT A Distribution Unlimited: PB Public Release			
13. SUPPLEMENTARY NOTES			
14. ABSTRACT Delegating tasks to autonomous machines and using them to mediate work interactions can offer many benefits, but at what cost? In human organizations, delegating tasks to others increases moral distance from the consequences of one's actions, but it is unclear whether this and related effects will carry over when delegating to machines. Will operating through an autonomous robot further undermine trust, increase risk-taking, reduce vigilance to threats and increase dehumanization of others? Or might it soften or even reverse these effects? What are the implications for power dynamics between humans when inserting autonomous machine representatives into the organizational chain of command? What methods of accountability are most effective when some decision-makers are machines? Social science research is needed to examine the psychological, organizational and cultural impact of these technological advances. With this grant, we are particularly interested in autonomy-mediated interactions. While technology has long allowed people to manage work through technology (e.g. email; videoconferencing), advances in AI allow algorithms to assume many of these managerial functions. These advances present a psychologically, socially, and organizationally unique situation: how do people respond to machines that act on the behalf of leaders or other teammates? This proposal brings together a multi-disciplinary team of experts to investigate how this increasing reliance on autonomy shapes individual and organizational decisions.			
15. SUBJECT TERMS			
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU 7
19a. NAME OF RESPONSIBLE PERSON LAURA STECKMAN			19b. PHONE NUMBER (Include area code) 426-7556

Standard Form 298 (Rev. 5/2020)
Prescribed by ANSI Std. Z39.18

Grant Number: AFOSR FA9550-18-1-0182

Project Title: Organizational Implications of Autonomy-Mediated Interaction

Start Date of Project: August 1, 2018

ICT Technical Point of Contact: Jonathan Gratch

Final Report

Proposal Title: Organizational Implications of Autonomy-Mediated Interaction

Technical Point of Contact: Jonathan Gratch, 310-440-0306, gratch@ict.usc.edu

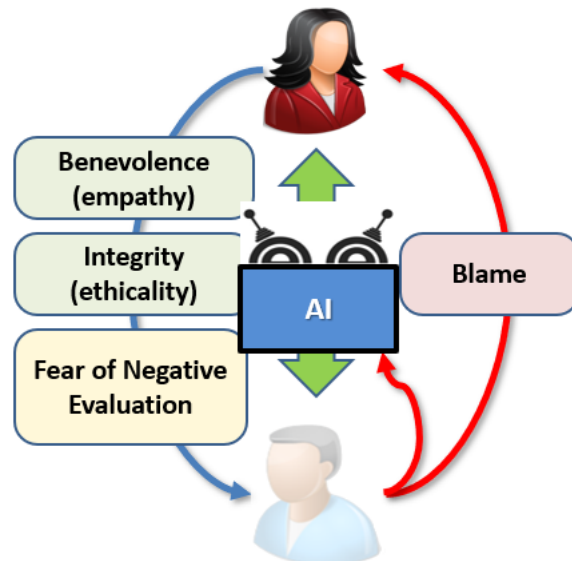
Date: October 26, 2023

Abstract: Delegating tasks to autonomous machines and using them to mediate work interactions can offer many benefits, but at what cost? In human organizations, delegating tasks to others increases moral distance from the consequences of one's actions, but it is unclear whether this and related effects will carry over when delegating to machines. Will operating through an autonomous robot further undermine trust, increase risk-taking, reduce vigilance to threats and increase dehumanization of others? Or might it soften or even reverse these effects? What are the implications for power dynamics between humans when inserting autonomous machine representatives into the organizational chain of command? What methods of accountability are most effective when some decision-makers are machines? Social science research is needed to examine the psychological, organizational and cultural impact of these technological advances. With this grant, we are particularly interested in *autonomy-mediated interactions*. While technology has long allowed people to manage work through technology (e.g. email; videoconferencing), advances in AI allow algorithms to assume many of these managerial functions. These advances present a psychologically, socially, and organizationally unique situation: how do people respond to machines that act on the behalf of leaders or other teammates? This proposal brings together a multi-disciplinary team of experts to investigate how this increasing reliance on autonomy shapes individual and organizational decisions.

Accomplishments

Research Objectives: The project examined “autonomy-mediated interaction” – meaning when people interact *via* an intelligent and autonomous system rather than via more traditional forms of computer-mediated communication (e.g., email or chat). An example of this would be delegating some organizational tasks and decisions to an intelligent assistant, or “algorithmic management” where a manager gives input to an algorithm that then oversees organizational functions. Whereas prior work on autonomous systems has tended to focus on how people interact with the system, here we focus on how people interact *through* the system. For example, we consider if people exert power over others differently when this power is mediated via autonomy, and whether the recipients of this power view the actions differently than if they were directly provided by a person.

Detailed Accomplishments: To investigate these questions, we performed behavioral experiments where human participants interacted with autonomous systems across a variety of domains and adopting different roles within these domains (e.g., is one exercising power over others, or is one the recipient of this power). Besides documenting differences in behavior and perceptions, these studies were designed to develop a theoretical model that helps explain how autonomy-mediated interaction differs from other forms of human social interaction. This model, and several of the supporting studies, were outlined in a review article published near the end of the project (Gratch & Fast, 2022) and outlined in the figure below.



Thus, studies were organized around several potential factors claimed to mediate how people exercise power through autonomy. First, we hypothesized that people would feel less empathy and, thus, act with less benevolence and integrity towards others when tasking their AI to perform social tasks. This was confirmed in several domains including standard economic games and negotiation tasks. For example, when offered an agent that could negotiate with people on their behalf, people were more willing to instruct the agent to act deceptively in order to gain an advantage in the negotiation.

Second, we hypothesized that autonomous systems would help deflect blame for unethical acts. In other words, rather than blaming the person that instructed the agent, people would tend to blame the AI. This was confirmed across several studies.

Finally, we hypothesized that people would anticipate that they would receive less blame for their negative actions, thus freeing them to engage in unethical behavior with more impunity. Again, this was confirmed in several studies.

Dissemination: Results of the research were disseminated through scientific publications in computer science, communication science and management venues. Results were also disseminated through a number of invited keynotes and other invited talks and through press interviews. A list of publications and talks is below:

Grant related journal publications:

- Arthur Jago, Roshni Raveendhran, Nathanael Fast and Jonathan Gratch. Algorithmic Management Diminishes Status: An Unintended Consequence of Using Machines to Perform Social Roles. *Journal of Experimental Social Psychology* (in press)
- Gratch, Jonathan, and Nathanael J. Fast. "The power to harm: AI assistants pave the way to unethical behavior." *Current Opinion in Psychology* 47. 2022
- Aike Horstmann, Jonathan Gratch and Nicole C. Kramer. I just wanna blame somebody, not something! Reactions to a computer agent giving negative feedback based on the instructions of a person. *International Journal of Human-Computer Studies* 154, 2021
- Daniel Dukes, Kathryn Abrams, Ralph Adolphs, et al. The Rise of Affectivism. *Nature Human Behavior* 5, 2021, p 816-820
- Celso de Melo, Jonathan Gratch and Frank Krueger. Heuristic thinking and altruism towards machines in people impacted by Covid-19. *iScience* 24 (3). 2021
- Jonathan Gratch. The Promise and Peril of Automated Negotiators. *Negotiation Journal* 37 (1). 2021, pp. 13-34
- Jonathan Gratch. The field of Affective Computing: An interdisciplinary perspective. *Transactions of the Japanese Society for Artificial Intelligence* 36(1), 2021, p. 4-12 (in Japanese)
- Minha Lee, Gale Lucas and Jonathan Gratch. Comparing mind perception in strategic exchanges: Human-agent negotiation, dictator and ultimatum games. *Journal on Multimodal User Interfaces* 15 (2), 2021, pp. 201-214
- Celso de Melo, Stacy Marsella and Jonathan Gratch. Risk of Injury in Moral Dilemmas With Autonomous Vehicles. *Frontiers in Robotics and AI* vol 7, 2021
- Johnathan Mell, Gale Lucas, Sharon Mozgai, Jonathan Gratch. The Effects of Experience on Deception in Human-Agent Negotiation. *Journal of Artificial Intelligence Research*. 2020
- Fast, N. J., & Jago, A. S. Privacy matters... Or does it? Algorithms, rationalization, and the erosion of concern for privacy. *Current opinion in psychology*, 31, 2020. pp 44-48
- Fast, N. J., & Schroeder, J. Power and decision making: new directions for research in the age of artificial intelligence. *Current opinion in psychology*, 33, 2020, pp. 172-176.
- Newman, D. T., Fast, N. J., & Harmon, D. J. When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*. 2020

Peer-reviewed Conference Publications

- Johnathan Mell, Gale Lucas and Jonathan Gratch. Varied Magnitude Favor Exchange in Human-Agent Negotiation. *20th ACM International Conference on Intelligent Virtual Agents*. Glasgow, Scotland 2020
- Reyhan Aydoğan, Tim Baarslag, Katsuhide Fujita, Johnathan Mell, Jonathan Gratch, Dave De Jonge, Yasser Mohammad, Shinji Nakadai, Satoshi Morinaga, Hirotaka Osawa, Claus Aranha and Catholijn Jonker. Research Challenges for the Automated Negotiating Agents Competition (ANAC) 2019. *7th International Conference on Agreement Technologies*. Greece 2020

- Yuushi Toyoda, Gale Lucas, Jonathan Gratch. The effects of autonomy and task meaning in algorithmic management of crowdwork. *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems*. Auckland, New Zealand, May 2020.
- Minha Lee, Gale Lucas, Johnathan Mell, Emmanuel Johnson and Jonathan Gratch. What's on Your Virtual Mind? Mind Perception in Human-Agent Negotiations. *19th International Conference on Intelligent Virtual Agents*. Paris, France, 2019
- Kerstin S. Haring, Jessica Tobias, Justin Waligora, Elizabeth Phillips, Nathan L. Tenhundfeld, Gale Lucas, Ewart J. de Visser, Jonathan Gratch and Chad Tossell. Conflict Mediation in Human-Machine Teaming: Using a Virtual Agent to Support Mission Planning and Debriefing. *The IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. New Delhi, India, October 2019
- Raveendhran, R., Jago, A., Fast, N.J., & Gratch, J. (August 2020). Voice Solicitation Through Technology. *Academy of Management Annual Conference*, (virtual 2020

Grant related invited talks:

- Keynote Speaker, HumanTech Summit, Warsaw, Poland, December 2023
- Invited Speaker, Behavioral Studies Colloquium. ETH Zurich. September 2023
- Keynote Speaker, ANIMATAS Symposium on Human-Machine Interaction, June 2022
- Invited Speaker, British Royal Society meeting on the Science of Face-to-face Interaction, April 2022
- Invited Speaker, University of Florida AI Seminar Series, November 2021
- Keynote Speaker, ANIMATAS Symposium on Human-Machine Interaction, October 2021
- Distinguished Lecture Speaker, UCSB Department of Communication, October 2021
- Invited Panel Chair, 9th International Conference on Affective Computing and Intelligent Interaction, September 2021
- Invited Speaker (Gratch), SurveyMonkey ML External Speaker Series, June 2021
- Keynote Speaker (Gratch), Conference on the Fantastic in Media Entertainment, May 2021
- Invited Seminar (Gratch), Max-Planck Center for Humans and Machines, November 2020
- Distinguished Speaker (Gratch), ISRE Webinar Series, October 2020
- Invited Panelist (Gratch), Harvard Program on Negotiation, Boston, May 2020
- Invited Speaker (Jago): Conference on Changing Nature of Work, Stanford, 2021
- Invited Seminar (Fast), McDonough School of Business, Management Department, Georgetown University, April, 2021.
- Invited Seminar (Fast), Viterbi School of Engineering, Civil and Environmental Engineering Department, University of Southern California, April 2021.
- Invited Seminar (Fast), Max-Planck Center for Humans and Machines, April, 2021
- Distinguished Speaker (Gratch), ISRE Webinar Series, October 2020
- Invited Panelist (Gratch), Harvard Program on Negotiation, Boston, May 2020
- Keynote Speaker (Gratch), 7th International Conference on Human-Agent Interaction, Kyoto Japan, October 2019

- Keynote Speaker (Gratch), Workshop on Methodology and/of Evaluation of IVAs, Paris, France, July 2019
- Invited talk, Cognitive Science Research Colloquium (Gratch), Northwestern University, March 2019
- Invited talk (Fast), Joint Colloquium, College of Engineering and Psychology Department, UC Santa Barbara, May, 2019
- Invited talk (Fast), Kenan-Flagler Business School, Organizational Behavior Department, University of North Carolina at Chapel Hill, January, 2019
- Invited talk (Fast), Anderson School of Management, Management and Organization Department, UCLA, February 2019
- Invited talk (Fast), Ross School of Business, University of Michigan, Management Department, October, 2018
- Invited talk (Fast), Haas School of Business, Management and Organizations Group, UC Berkeley, September, 2018

Development of the principal discipline(s) of the project:

The research was a collaboration between the disciplines of computer science and psychology, addressing key issues around the design of intelligent teammates and how design choices impact user and organizational behavior. Thus, the primary contributions centered around the development of theoretical frameworks that could promote effective and ethical AI systems. Perhaps the chief contribution is highlight a larger context when assessing the effectiveness of human-machine interaction. Most prior work on the impact of autonomy has focused on how people interact *with* intelligent machines, but has not considered how people interact *through* intelligent machines. With the rise of intelligent assistants, people are increasingly empowered to delegate social tasks to these AI assistants. Thus, our work has highlighted the importance of critically examining how people choose to assign tasks to their digital assistants and the values implicit in this tasking.

Other disciplines:

Though arising out of a collaboration between computer science and psychology, the research has a significant impact on business research and the publications have been presented and published in business venues such as Academy of Management.

Impact on development of human resources

The project provided opportunities for research and skill development for several PhD students and PostDocs, including underrepresented groups, that are now contributing to U.S. research and education. Two PhD students completed their PhD dissertations with support from the project and are now assistant professors at US Universities – Emmanuel Johnson (African-American) accepted an Assistant Professor of Computer Science position at the University of Florida. Johnathan Mell is an Assistant Professor of Computer Science at the University of Central Florida. Two Postdoctoral students worked on the project and are now assistant professors – Roshni Raveendhran (Indian-American woman) is an Assistant Professor of Business Administration at the University of Virginia and Arthur Jago is an Assistant Professor of Management at the University of Washington, Tacoma.

Additionally, the project supported the PhD training of two students from underrepresented groups (Jessie Hoegen and Su Lei) and one additional PhD student. The work supported several research internships.

Staff supported under grant: The following personal were supported under the period of performance of the grant this period

- Faculty:
 - Jonathan Gratch (USC Computer Science & Psychology)
 - Nathanael Fast (USC, Business)
- Postdoctoral Researcher
 - Oliver Fisher (USC)
 - Arthur Jago (USC)
 - Joo-Wha Hong (USC)
- PhD students
 - Emmanuel Johnson (USC)
 - James Hale (USC)
 - Su Lei (USC)
 - Jessie Hoegen (USC)
 - Merrick Osborne (USC)
 - Maya Cratsley (USC)
 - Johnathan Mell (USC)
- Undergrad Students
 - Chatali Bharat Joshi (USC)
 - Hride Patel (USC)
- Visiting Graduate Students
 - Aike Horstmann (University of Duisburg-Essen)
 - Yixian Li (Western University, Canada)
- Research Staff
 - Alesia Gainer
 - Elizabeth QUinn
- External (unfunded) collaborators
 - Roshni Raveendhran (University of Virginia)
 - Arthur Jago (University of Washington, Tacoma)
 - Celso de Melo (Army Research Laboratory)