

Emergent Behaviors in Human-Robot Systems

RSS 2020 Workshop

July 12, 2020

Oregon State University at Corvallis, Oregon, USA Virtual Event

Key Facts

Web: iliad.stanford.edu/rss-workshop **Date:** July 12, 2020 (Full Day) **Contact:** ebiyik@stanford.edu

Confirmed Speakers



[Brenna Argall](#)
Northwestern University



[Anca Dragan](#)
UC Berkeley



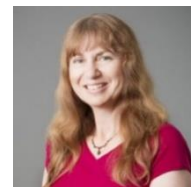
[Judith Fan](#)
UC San Diego



[Jakob Foerster](#)
Facebook AI &
University of Toronto



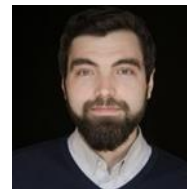
[Robert D. Hawkins](#)
Princeton University



[Maja Mataric](#)
University of
Southern California



[Negar Mehr](#)
Stanford University &
University of Illinois
Urbana-Champaign



[Igor Mordatch](#)
Google Brain



[Harold Soh](#)
National University
of Singapore

Abstract

Robots are increasingly becoming members of our everyday community. Self-driving cars, robot teams, and social and assistive robots operate alongside human end-users to carry out various tasks. Similar to the conventions between humans, which are the low dimensional representations that capture the interaction and can change over time, emergent behaviors form as a result of repeated long-term interactions in multi-agent systems or human-robot teams. Unfortunately, these emergent behaviors are still not well understood. For instance, the robotics community has observed that many different, and often surprising, robot behaviors can emerge when robots are equipped with artificial intelligence and machine learning techniques. While some of these emergent behaviors are simply undesirable side-effects of misspecified objectives, many of them significantly contribute to the performance in the task and influence other agents in the environment. These behaviors can further lead to developing conventions and adaptation of other agents, who are possibly humans, by encouraging them to approach the task differently.

Goal. We want to investigate how complex and/or unexpected robot behaviors emerge in human-robot systems, and to understand how we can minimize their risks and maximize their benefits. This workshop promotes a discussion on

- How emergent behaviors form in human-robot, human-human, and robot-robot settings
- How we can predict emergent behaviors
- What types of emergent behaviors we would like to see in real-world applications
- How their negative consequences and risks might be alleviated
- How they can be better utilized for more efficient human-robot or multi-agent interaction

Organizing Committee

Erdem Biyik
Noah Goodman

Minae Kwon
Stefanos Nikolaidis

Dylan Losey
Dorsa Sadigh