

Using Gesture and Language to Establish Multimodal Conventions in Collaborative Physical Tasks

Kiyosu Maeda

Princeton University, Princeton, New Jersey, United States

Ching-Yi Tsai

Princeton University, Princeton, New Jersey, United States

Judith Fan

Stanford University, Stanford, California, United States

Parastoo Abtahi

Princeton University, Princeton, New Jersey, United States

Abstract

A quintessential feature of human intelligence is the ability to create ad-hoc conventions over time to accomplish shared goals efficiently. Prior research has primarily focused on unimodal communication or communication mediated by a 2D screen. We study how multimodal communication using gesture and language changes during physical collaboration. We paired human participants and used augmented reality to isolate voice and hand gestures. One participant saw a 3D virtual tower and provided instructions to the other participant, who constructed the physical tower. Participants became faster and more accurate by forming conventions that made use of both gestural and linguistic abstractions. Redundancy was used to emphasize a change in the established convention. Based on these findings, we extend unimodal probabilistic models of convention formation to multimodal settings while capturing various modality preferences. Our work serves as a building block for convention-aware and physically situated intelligent agents.