Human-AI Cooperation: Leveraging AI Norm Reinforcement in a Public Goods Game

Nico Mutzner
University of Zurich

Taha Yasseri
Trinity College Dublin

Heiko Rauhut
University of Zurich

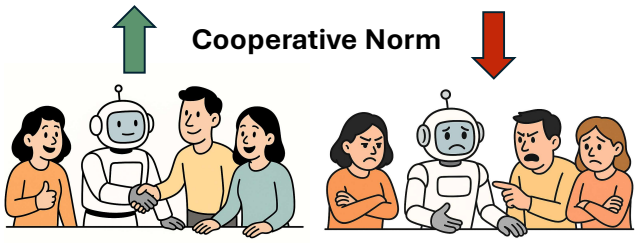
SOHAM
Centre for Sociology of Humans and Machines

Universität Zürich[UZH]

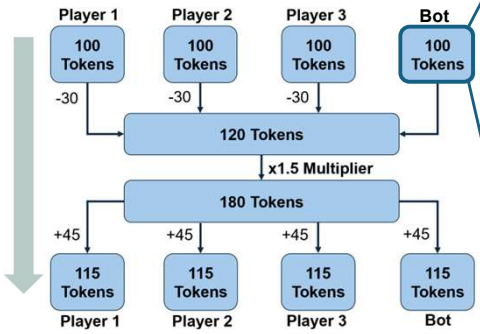
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Research Question

How does the implementation of AI agents reinforce or erode cooperative norms in group settings?



Public Goods Game



2x3 Design

Human vs AI

x

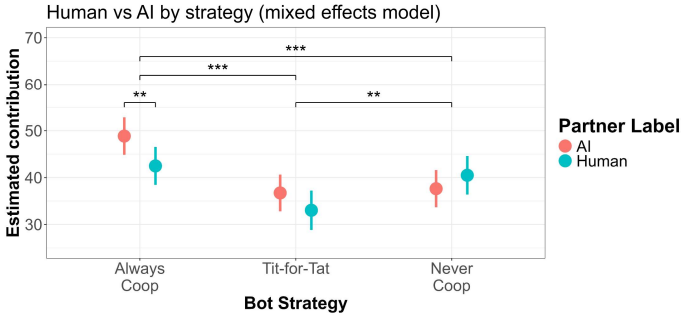
- Always coop
- Tit-for-tat
- Never coop

Rand. Control Trial
Participants 238

Our experiment shows **AI teammates can spark as much and sometimes more cooperation** than humans

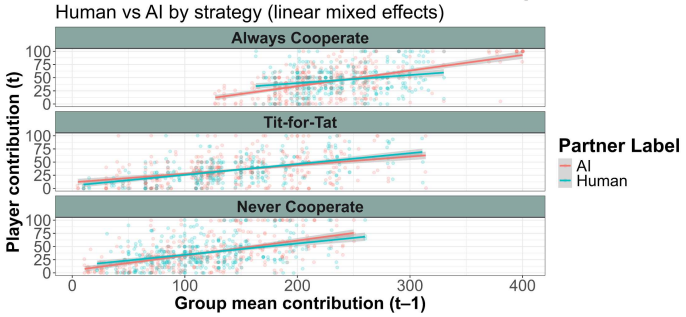
Estimated Contributions in Public Goods Game

Human vs AI by strategy (mixed effects model)



Round Contribution vs. Last-Round Group Mean

Human vs AI by strategy (linear mixed effects)



- **Algorithm Preference:** AI partners match human partners in most cases, yet unconditional cooperative AI triggers increased cooperation.
- **Go big or go rogue:** Unconditional cooperators inspire cooperation, while Tit-for-Tat elicits less than non-cooperators.
- **Group Influence:** People normatively adjust to prior group contributions across labels, with strategy driving the strength.

Cooperation follows actions. Design AI for strategy, not identity.

Contact:
Nico Mutzner
Nico.Mutzner@uzh.ch

SCAN ME