Coordination and Cooperation: Trusting Autonomous Agents

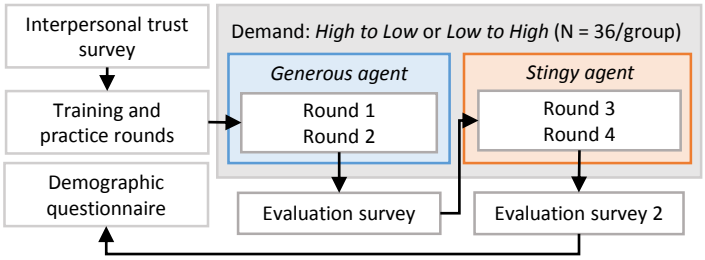
Erin K. Chiou, John D. Lee, Anousone K. Bounket, Qiwen Shu, Kim Le

Trust in automation: A reappraisal

Advances in computational and networked technologies have resulted in highly capable machines. These machines have the potential to move beyond their roles as subordinates and resemble autonomous partners, interacting with human teammates, and coordinating actions in a changing environment[1]. To reach the next level of systems safety in high criticality domains, these human-automation teams must be able to adapt in unpredictable environments, and avoid becoming brittle systems that fail when unexpected events occur. The potential failure modes of human-machine systems have been studied through reliance on and compliance with technology, both of which are mediated by trust[2]. With these new machine capabilities, this project focuses rather on coordination and cooperation with autonomous machine agents, and the interdependence of team interactions.

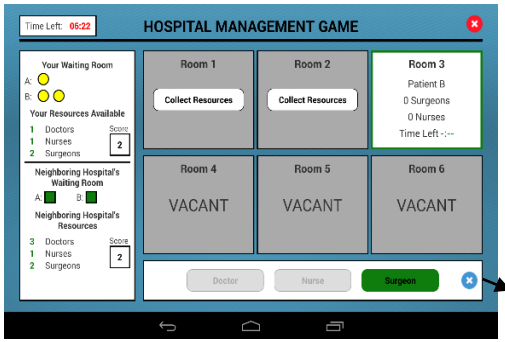
Study conditions and procedure

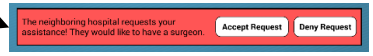
- Environment: high to low demand, low to high demand.
- Agent behavior: high or low cooperation, order of exposure.
  - High coop: requests if a participant has it, accepts requests most of the time.
  - Low coop: requests when it needs it, accepts requests some of the time.



Microworld environment

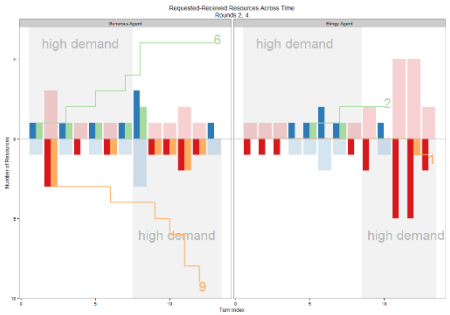
- Developed using Android SDK and played in Genymotion using a 15” laptop.
- Participants played one demand condition, agents played the complementary one.
- Players needed to request resources and accept requests to achieve optimal scores.

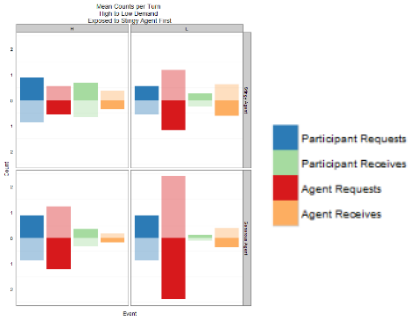




Preliminary results

To visualize interdependence, we plotted player actions across time. Below left is a participant in the High to Low demand condition, exposed to the Stingy agent first. Below right are the mean counts of participants in the same condition.





We used an affinity diagram to cluster participants into behavior profiles. These profiles were then distilled into five model parameters:

Stinginess – willingness to give
Greed – timing of requests
Curiosity – tests responses
Flexibility – adapts to conditions
Awareness – sensitivity to mechanics

Future work will include sequential analysis of the interactions to determine emergent patterns of cooperative decision making, and comparison to discrete event simulation.

[1] Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–85.
[2] Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. Human Factors, 46(1), 50–80.
*This work was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256259.