Romance of the Three Kingdoms

Final HRI Project Presentation

Akanksha Saran
Benjamin Choi
Ronald Lai
Wentao Liu
Contents

- Project Recap
- Experimental Setup
- Results and Discussion
- Conclusion
Project Recap

- Find if humans trust robots or other humans more
- ...via a 3-player game.
  - Robot, a human confederate, and the participant.
  - Two game policies: trustworthy and untrustworthy.
  - Robot and confederate always follow opposing gameplay policies.
Experimental Setup

- Conference room in Gates
- 20 participants, mostly technical majors and friends of ours.
- Data collection:
  - Logs of in-game behaviors.
  - Post-game survey (Google Form)
Video Time!
You are connected as player 0

Round 3 of 8

6

- Attack player 1
- Attack player 2

Player 1

Player 1: Ally with me?
You: No

Player 2

You: Ally with me?
Player 2: Yes
You: 😊
Player 2: That's deep man
Player 2: Ally with me?
You: No
Player 2: 😞
Player 2: Ally with me?

Your total score: -2

<table>
<thead>
<tr>
<th>Round</th>
<th>Change to score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Send

Send
Results
Participant Pool

- 20 participants, ages 18-23 yrs.
- 15 males, 5 females
- 14 pet-owners,
  6 non-pet owners

Policy 1 - Untrustworthy Robot, Trustworthy Confederate
Policy 2 - Trustworthy Robot, Untrustworthy Human
Results

In-game behaviors
Results: Probability of Attacking Robot

- Overall, participants are more likely to attack robot than human confederate.
- Also, probability to attack robot is higher when robot is running untrustworthy policy.
- In the first round, 75% of participants attack the robot.
Results: Probability of Attacking Robot

Examining attacks in the 8th (final) round of each game.

Not as expected, the trustworthy robot was attacked more often in the final round.

Two possible explanations:
1) Either participants didn’t gauge the right policy due to the randomness in the game
2) Participants were playing to win and flipping the attacks to maximize their own scores
Betrayal

- Untrustworthy robots and confederates are more likely to be backstabbed

- Untrustworthy robots most likely to be backstabbed

![Bar chart showing probability of being backstabbed for robots and humans. The chart indicates a higher probability for robots compared to humans, with the category 'Untrustworthy' showing significantly higher values than 'Trustworthy'.]
Results

Survey responses
Results: Trust in confederate correlated with policy

- Participant reported more trust in confederate when confederate is following trustworthy policy.
- However, the same is not true for reported trust in robot (weak statistical correlation between reported trust and policy of robot).
Results: Trust in confederate correlated with probability of attacking robot
Pet-owners v/s Non-pet owners

- No significant difference in comforts with robots or willingness to play again
- Pet-owners trust in the robot moderately higher than non-pet owners
- Pet owners like the robot significantly more
Introverts v/s Extroverts

- Extroverts showed more willingness to interact again with the robot
Willingness to play again*

- More willingness to play again with a robot than human player
  - Excitement and interest towards the robot?

*Open-ended question: 19/20 participants answered ‘Yes/Sure’ or ‘No’
Conclusions & Limitations

- People tend to be more comfortable attacking the robot than the plant.
  - Interested in figuring out the robot opponent’s strategy?
    - “How does robot play?”
    - “…I'd like to know more about how the robot decided to attack.”

- More willing to ‘offend’ robot than person

- Conduct more rounds per player to see if this effect persists?
Conclusions & Limitations

- Limitations of our point system.
  - People seemed to be flipping their decisions towards the end in the end. Attempting to ‘even out’ the score?
    - Participants were playing to win, overshadowing the trust factor
    - Concept of “fairness”? Participant feel bad attacking the same player throughout the game?
Conclusions & Limitations

- Differences observed between 2 sets of policies not significant
  - The robot might have not allied with the player enough to give him/her a sense of its policy
  - Inherently chaotic situation in each game may have resulted in undesired randomness
  - Policies could have been more extreme to reduce the inherent randomness in the game.
Conclusions & Limitations

- Participants tended to look at the screen throughout the game, very little attention given to physical robot despite its embodiment
  - Screen as the primary means (and hence focus) of engagement
  - Embodiment of robot and means of interacting with robot are separated to some degree.
  - Perceived differentiation between human and robot opponents is reduced.
  - Modifications to game to force embodied interaction?
Thanks!