Assessing the changes in human trustworthiness as a result of an artificial agent directing the human in joint activities

Iulia - Nicoleta Dinu I.N.Dinu@student.tudelft.nl
Supervisor: Carolina Jorge Responsible Professor: Dr. Myrthe Tielman

Background

Human Trustworthiness = The human's quality of being worthy to be trusted

Directing = Giving commands regarding the next move that the human should make, however the human is still autonomous

Joint activity = Activity that requires collaboration between human(s) and artificial agent(s)

High Interdependence = The two parties are required to collaborate in order to finish the game (i.e. one cannot win without the other)

- Joint activity => risk
- Trust is most relevant if there is risk
- Trustworthiness is about more than capability
- Until now, the emphasis on the trustworthiness of the artificial agent from human's perspective

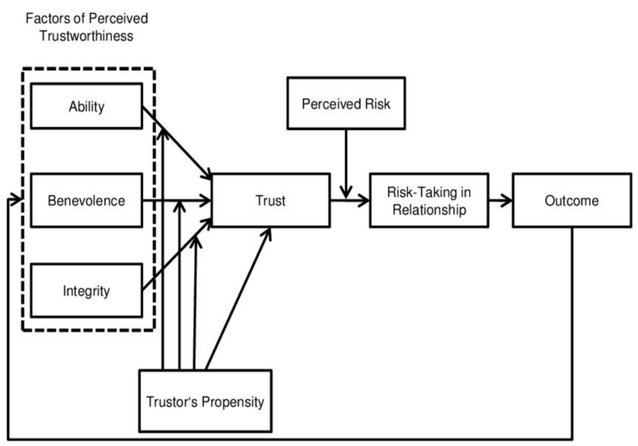


Figure 1:
Trust Model based on Mayer, Davis and Schoorman (1995)

ABI trust model

- Ability
  - Are you able given the circumstances?
- Benevolence
  - Interpersonal relationship
  - To which extent are you believed to want to do good to the trustor?
- Integrity
  - Shared personal values
  - Honest, fair, principled

Hypothesis

The human trustworthiness will increase when the artificial agent is directing the human in joint activities.

Methodology

Experimental Setup

- 20 participants for the control group
- 20 participants for the experimental group
- Search and Rescue Game
- MATRX software
- Directing agent was implemented for the experimental group

Quantifying trustworthiness

- Objective measurements
  - No. of ticks, Amount of lies, etc.
- Subjective measurements
  - 15 Likert scale questions

Tests and Measurements

- Mean & Standard Deviation
- Shapiro-Wilk test => normality of the data
- Cronbach's alpha => internal consistency
- Mann–Whitney U test
- Two-tailed test

Reference

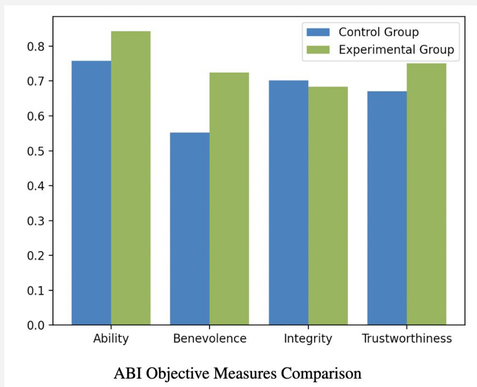
Towards a Formative Measurement Model for Trust - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Trust-Model-based-on-Mayer-Davis-and-Schoorman-1995_fig1_228798197 [accessed 14 May, 2022]

Figure 2:
Search and Rescue game in MATRX

Results & Discussion

Objective measurements

- Mann–Whitney U test
  - Ability → **not** significant
    - Expected due to the strong background in Computer Science
  - Integrity → **not** significant
- Two-tailed test
  - Benevolence → significant
    - t(38) = -2.673, p = .011
    - Obedience to authority = tendency to comply with authority figures
  - Trustworthiness → **not** significant

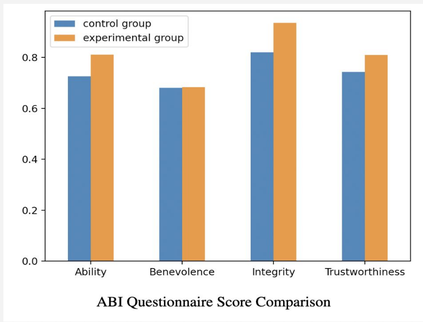


ABI Objective Measures Comparison

Discussion

Subjective measurements

- Mann–Whitney U test
  - Integrity → significant
    - $U(N_{control} = 20, N_{experimental} = 20) = -2.257, p = .03$
    - Obedience to authority
- Two-tailed test
  - Ability → significant
    - Cronbach's alpha for experimental group = 0.504 => low consistency
    - t(38) = -2.257, p = .03
  - Benevolence → **not** significant
  - Trustworthiness → **not** significant



ABI Questionnaire Score Comparison

Limitations

- Small sample size
- Lack of diversity
- English proficiency
- Computer Science Background

Future Work

- Repeat the experiment more participants with various backgrounds
  - VoiceOver feature
- Optimise artificial agent's implementation

TU Delft