The effects of an agent asking for help on human’s trustworthiness in human-AI teams

Obame Obiang Christopher Armel (C.A.ObameObiang@student.tudelft.nl) – Supervised By Carolina Ferreira Gomes Centeio Jorge, Myrthe Tielman

TU Delft Delft University of Technology

1.Background

- ❑ Human-AI teams are more efficient than autonomous robots
- ❑ Mutual Trust is needed for an efficient collaboration [1]
- ❑ For an agent A to trust an agent B, agent B must appear trustworthy to A
- ❑ **Trustworthiness** : “a trustworthy entity presents high values of competence, integrity, benevolence, and predictability in the situation in assessment” [2]

2.Problem Analysis

- ❑ There is a lack of knowledge concerning the trust of an artificial agent towards its human partner
- ❑ Investigating how the artificial agent's behaviour may influence the human's trustworthiness would give more knowledge on the topic

3.Research Question

- ❑ *How does an artificial agent asking human for advice/help affect human trustworthiness?*
- ❑ **Hypothesis** : The trustworthiness of the human increases when the artificial agent ask him for help or advice.

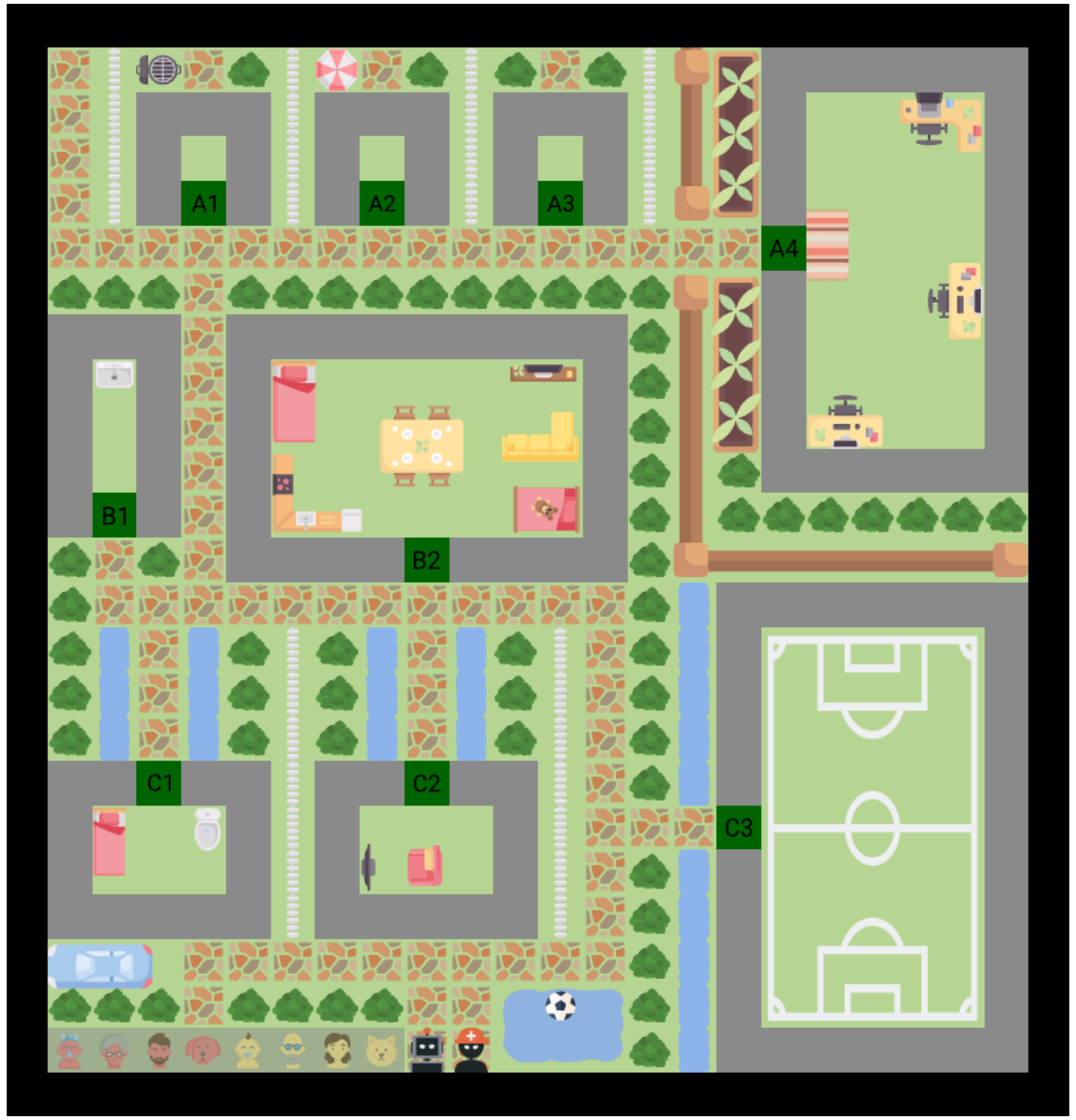

Figure 1 : Urban Search and Rescue MATRX

4.Methodology

Controlled Experiment

- ❑ 1 human and 1 agent collaborating
- ❑ Task : Rescuing victims alongside an artificial agent
- ❑ Participants : 28
- ❑ Control Group : Normal Agent
- ❑ Experimental Group : help-seeker Agent

Analysis

- ❑ Objective Measures: Game completion time, number of victims rescued, number of messages sent….
- ❑ Subjective Measures : Questionnaire
- ❑ If results are normally distributed, an independent t-test will be used
- ❑ If not, the independent Mann-Whitney test will be used

Help-seeker agent Design

- ❑ Will ask advice on which room to search next
- ❑ Will ask for help when searching the larger rooms
- ❑ Will ask advice as to which found victim to rescue next

5.Results

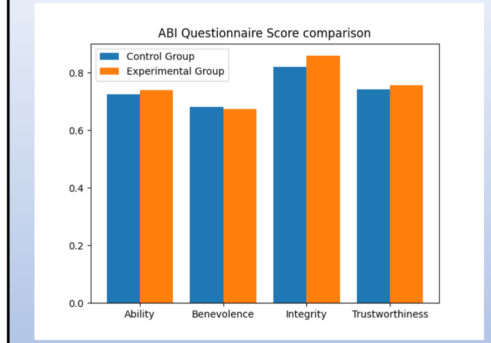

Figure 2 : ABI subjective measure comparison

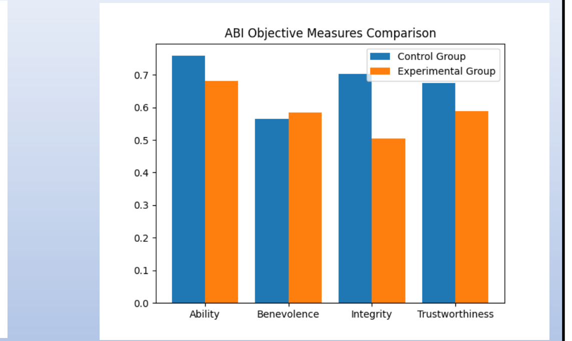

Figure 3 : ABI objective measure comparison

- ❑ No significant results were found
- ❑ Objective trustworthiness : t(26) = 1.19, p = 0.25
- ❑ Subjective trustworthiness : t(26) = 0.219, p = 0.83

6.Conclusion and Future Work

- ❑ The results of the experiment are inconclusive.
- ❑ The lack of significant results from the subjective measurements may be due to a bias from the human participant.
- ❑ Testing the effect of asking for help using a different trust model should be considered For future Work.
- ❑ Future work may consider using different contexts to test the interaction between the agent and the human, for longer periods of time.

[1] E. Salas, D. E. Sims, and C. S. Burke, “Is there a Big Five in Teamwork?” Small Group Research, vol. 36, no. 5, pp. 555–599, 10 2005. [Online]. Available : https://doi.org/10.1177/1046496405277134, [2] Joana Urbano, Ana Paula Rocha, and Eugénio Oliveira. A Socio-cognitive Perspective of Trust. In Agreement Technologies, pages 419–429. Springer Netherlands, Dordrecht, 2013.