HHAI2022: Augmenting Human Intellect
S. Schlobach et al. (Eds.)

doi:10.3233/FAIA220214

Using a Virtual Reality House-Search Task to Measure Trust During Human-Agent Interaction (Demo Paper)

Esther KOX[a,b,1], Jonathan BARNHOORN[a] , Lucía RÁBAGO MAYER[b], Arda TEMEL[b], Tessa KLUNDER[a]
[a] *TNO, Soesterberg, The Netherlands*
[b] *University of Twente, Enschede, The Netherlands*

Abstract. How can we observe how people respond to consequential errors by an artificial agent in a realistic yet highly controllable environment? We created a threat-detection house-search task in virtual reality in which participants form a Human-Agent Team (HAT) with an autonomous drone. By simulating risk, we amplify the feeling of reliance and the importance of trust in the agent. This paradigm allows for ecologically valid research that provides more insight into crucial human-agent team dynamics such as trust and situational awareness.

Keywords. Virtual Reality; Interaction paradigm, Human-Agent Collaboration; Teamwork; Trust

1. Introduction

Trust is a fundamental aspect of any form of teamwork. Definitions of trust often contain the willingness to be vulnerable in a context of risk. So in order to understand to what extent humans are willing to trust an artificial agent (AA) in circumstances characterized by threat, we need realistic research environments involving risk [1]. Prior studies studying Human-Agent trust involve relatively low-risk tasks [2]–[4], as it is ethically and practically challenging to let participants truly experience risk in a research environment. To overcome this problem, we have developed a virtual reality (VR) application that simulates a hazardous environment where participants feel threatened, without ever being in actual danger. The scene is realistic yet highly controllable and allows us to directly observe behaviour. With this unique VR environment we study how people respond to consequential errors by an AA, which provides us with more insight into essential HAT dynamics such as the development of trust and situational awareness.

2. VR Environment

The VR environment resembles a military threat-detection house-search task (Fig 1). It consists of two buildings, each with three floors. The two buildings are designed to be similar, but include different details. While searching the buildings, participants are accompanied by a drone that tells them whether it detects threats in the environment by advising them to either move carefully or to proceed normally via automated audio

[1] Esther Kox, E-mail: esther.kox@tno.nl, e.s.kox@utwente.nl

messages (see [5]). Level of trust in the drone is repeatedly measured using a virtual slider in the VR environment [4], [6].

During the search, participants encounter multiple obstacles, such as a laser trap and a safety ribbon. The agent provides the instructions on how to dismantle the trap (by cutting a wire) or how to cut the safety ribbon. These obstacles and associated actions are included to enhance the immersion. Participants also encounter a burglar and a smoking and beeping IED (improvised explosive device, Fig 2). These latter events were designed to startle the participant without harmful consequences; the burglar screams but then runs off and the IED smokes and beeps, but does not fully explode. The agents fails to warn participants for these events which allows us to examine the dynamics of trust following a violation in trust.

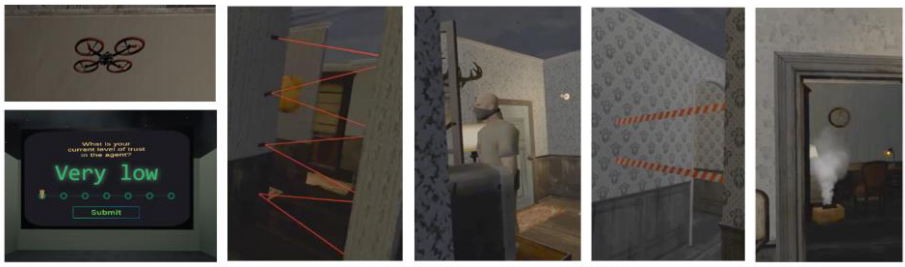

Figure 2. Screenshots of the VR environment.

All events in the virtual environment (e.g., audio messages, animations and transport to the next floor) are automated using invisible triggers placed in the virtual space. When the participant walks into such as a trigger, the corresponding event was activated.

The VR environment was built in Unity 3D (version 2020.4.3.F1). Participants used and Oculus Rift HMD (head-mounted device) and two hand controllers (Oculus Touch) to experience and interact with the environment. The Cyberith Virtualizer ELITE 2, a 360 degrees VR Treadmill with an implemented motion platform, allowed participants to walk in the virtual environment.

3. Conclusion

The VR environment offers ecological validity, experimental control, reproducibility [7], and emotional engagement of participants [8]. More than 2D screen videos or games, VR has the ability to create a strong sense of presence and to increase sympathetic activation significantly [9]. We suspect that the current VR environment intensifies feelings of risk and betrayal after a trust violation in comparison to previous studies that used 2D simulations or using videos [10]. This was often visible as a number of participants startled, flinched or cursed when the burglar or the bomb were encountered and as participants reported that they perceived the environment as threatening. These intensified feelings could be more representative of non-simulated human-AI interactions than 2D screen videos and games and thereby offer some important insights in the future of HATs. For upcoming studies we would like to include physiological and behavioural measures associated with trust, like heartrate and walking speed.

References

[1] G. Matthews, A. R. Panganiban, R. Bailey, and J. Lin, "Trust in Autonomous Systems for Threat Analysis: A Simulation Methodology," in *International Conference of Virtual, Augmented and Mixed Reality*, 2018, vol. 10910 LNCS, pp. 116–125, doi: 10.1007/978-3-319-91584-5_10.

[2] E. J. de Visser *et al.*, "Almost human: Anthropomorphism increases trust resilience in cognitive agents.," *J. Exp. Psychol. Appl.*, vol. 22, no. 3, pp. 331–349, Sep. 2016, doi: 10.1037/xap0000092.

[3] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, "Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction," *25th IEEE Int. Symp. Robot Hum. Interact. Commun. RO-MAN 2016*, pp. 493–500, 2016, doi: 10.1109/ROMAN.2016.7745163.

[4] T. Kim and H. Song, "How should intelligent agents apologize to restore trust ?: The interaction effect between anthropomorphism and apology attribution on trust repair," *Telemat. Informatics*, 2021.

[5] E. S. Kox, L. B. Siegling, and J. H. Kerstholt, "Trust development in military and civilian Human-Agent Teams: the effect of social-cognitive recovery strategies," *Int. J. Soc. Robot.*, 2022, doi: 10.1007/s12369-022-00871-4.

[6] P. W. de Vries, C. Midden, and D. Bouwhuis, "The effects of errors on system trust, self-confidence, and the allocation of control in route planning," *Int. J. Hum. Comput. Stud.*, vol. 58, no. 6, pp. 719–735, 2003, doi: 10.1016/S1071-5819(03)00039-9.

[7] X. Pan and A. F. d. C. Hamilton, "Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape," *Br. J. Psychol.*, vol. 109, no. 3, pp. 395–417, 2018, doi: 10.1111/bjop.12290.

[8] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences," *Front. Hum. Neurosci.*, vol. 9, no. DEC, pp. 1–19, 2015, doi: 10.3389/fnhum.2015.00660.

[9] A. Chirico, P. Cipresso, D. B. Yaden, F. Biassoni, G. Riva, and A. Gaggioli, "Effectiveness of Immersive Videos in Inducing Awe: An Experimental Study," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017, doi: 10.1038/s41598-017-01242-0.

[10] E. S. Kox, J. H. Kerstholt, T. F. Hueting, and P. W. de Vries, "Trust repair in human-agent teams: the effectiveness of explanations and expressing regret," *Auton. Agent. Multi. Agent. Syst.*, vol. 35, no. 2, pp. 1–20, 2021, doi: 10.1007/s10458-021-09515-9.