

Human-Agent Trust Relationships in a Real-Time Collaborative Game

Sylvain Daronnat
University of Strathclyde
Glasgow, Scotland, UK
sylvain.daronnat@strath.ac.uk

ABSTRACT

Collaborative virtual agents are often deployed to help users make decisions in real-time. For this collaboration to work, users must adequately trust the agents that they are interacting with. In my research, we use a game where human-agent interactions are recorded via a logging system and survey instruments in order to explore this trust relationships. We then study the impact that different agents have on reliance, performance, cognitive load and trust. We seek to understand which aspects of an agent influence the development of trust the most. We hope to pave the way for trust-aware agents, capable of adapting their behaviours to users in real-time.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *HCI design and evaluation methods*.

KEYWORDS

HCI, Trust, Reliance, Cognitive Load, Performance, Task Difficulty, Collaborative Game, Agents

ACM Reference Format:

Sylvain Daronnat. 2018. Human-Agent Trust Relationships in a Real-Time Collaborative Game. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

An increasing number of collaborative agents are being deployed in various areas, from safety critical tasks in the military and medical domains [3, 19, 20] to more leisure-oriented activities involving collaborative game and virtual companionship [5, 9, 24].

Our research focuses on the assessment of trust in real-time and the investigation of factors that influence its development the most. To this end, we created a human-agent collaborative game where the agents' behaviours, performance as well as the task difficulty and predictability are controlled. We log users' interactions in real-time using an integrated logging system and record their answers to pre and post task questionnaires. The framework created for this research was presented during the Everyday Automation Experience

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

CHI2019 Workshop [7]. Two lab-based studies have been conducted since then, the first one focusing on the impact of agents' reliability on players (currently under review) and the second one focusing on how agents' errors impact players' behaviours [6]. A third online-based study is currently under way. In this new study, we explore whether players are more likely to rely and trust an agent under different visual conditions affecting on-screen visibility.

2 RESEARCH BACKGROUND

Interactions with collaborative agents are often done within command and control scenarios, where users can choose to take into account or disregard the agent's input [3, 20]. Human-agent and human-human trust relationships share similarities as both help to achieve goals in an environment characterised by uncertainty and vulnerability [18]. Adequate trust in automation has been shown to have a positive impact on task performance and situational awareness [8, 10, 11, 25].

Most of the studies up to date were carried out in turn-based settings, where trust is assessed via pre- and post-hoc surveys [12, 21, 23]. However trust is a multi-modal concept influenced by numerous factors, which is continuously evolving over time and through interactions [14].

There is currently a lack of studies focusing on the empirical assessment of real-time trust relationships in human-agent interactions [1]. Gamification studies provide ideal environments to explore such relationships by mixing simple and accessible gameplay with an engaging task in a controlled environment [4]. To meet this need, we propose a human-agent interactive framework in which both quantitative and qualitative information is recorded and used to analyse the evolution of human-agent relationships overtime. We believe that through the analysis of objective (gathered through interaction) and subjective (gathered through survey instruments) data, we can elicit and later predict the evolution of reliance and trust in a collaborative agent.

3 RESEARCH QUESTIONS

We seek to understand how trust in agents evolves over time and how it relates to reliance, cognitive load, task performance and situational awareness. Our work focuses on the consequences of imperfect automation on users' behaviours and perception of collaborative agents. Namely, we focus on the following research questions:

- (1) How do changes in the predictability of an agent affect users' behaviours and perception of automation?
- (2) How do different types of agent errors affect users' perceptions of the agents?

- (3) Is trust in an agent affected by different levels of uncertainty in the environment?
- (4) Can an agent use different strategies to try and recover the users' trust after a failure?

4 RESEARCH APPROACH

To study the evolution of users' behaviours while collaborating with agents, we created an interactive task in the form of a video game where players have to collaborate with different agents in order to destroy waves of incoming missiles (see Figure 1.). The game was heavily influenced by Missile Command [15] thanks to its accessible gameplay that is ideal for user studies [4]. During gameplay, the player can allow the agent to take care of the aiming process, or can choose to override it. In all cases, the player is the only one able to fire a projectile, which reduces the potential for complacency [2] while keeping a minimum level of task engagement. Our controlled environment allows us to adjust the task complexity, agents behaviours and accuracy as well as visual uncertainty in order to test the resulting effects on players' behaviours and perception of the agents. An in-game logging system as well as survey instruments [13, 16] were incorporated in the game to collect relevant information such as real-time performance (how well the tasks' goals were met) and reliance (how long the player actually relied on the agent). The game was created using the open source Godot Game Engine [17].



Figure 1: Screen-capture of the game-framework used in our research.

5 CONTRIBUTION

So far, we have used the aforementioned game-framework to conduct three studies:

Testing Agents Reliability. In our first study (currently under review) we set out to explore the relationships between trust, agents' predictability and agents' reliability. We designed a lab-based within-groups experiment ($n = 30$) where players interacted with agents that displayed different levels of performance and/or reliability. We found that, at the same level of agents' performance, participants reported higher trust, lower cognitive loads and were more likely to rely on more predictable agents. However, the same effects were not visible as agents' performance decreased. These

findings further highlight the importance of predictability and consistency in the design of potentially error-prone agents, and how it impacts human-agent collaboration in real-time. We also found that it was possible to infer users' trust ratings with a high level of accuracy using both behavioural interaction data and qualitative assessment of the agents and task.

Testing Types of Agents Errors. In our second study [6] ($n = 24$), we tested the impact of different agents performing at the same level in terms of reliability but displaying different behaviours when it comes to making errors. Agents' behaviours and their associated types of errors were created using Reason's human errors taxonomy [22], which has been later contextualised in Human-Agent collaborative settings by Baker et al [1]. Our results showed that different types of agent errors have a different impact on users propensity to trust and rely on an agents' inputs despite the errors having a similar impact on performance. In most cases, we found that an error caused by the agent's inaction had a less negative impact on trust and performance compared to errors of planning or commission.

Impact of Visual Uncertainty on Trust. We are currently running our third study, where we look at the impact of visual uncertainty (partial and near-total) on participants' propensity to rely, trust and perform tasks while collaborating with agents. Due to the COVID pandemic, the study has been moved online. For now, preliminary results ($n = 12$) seem to indicate that users report higher trust in the agents when visibility is the most reduced even when they have previously interacted with the agent in higher visibility settings. However, situational awareness does not significantly change from situations with perfect visual information to situations where the visibility is reduced.

6 DISSERTATION STATUS

Relevant literature review as well as the design of the main component of our framework has been completed. Lab-based piloting sessions and science-fair events were used to calibrate the different settings of our framework (game mechanics, task complexity). Recently the study has been adapted to an online platform, where participants can take part in the study remotely, allowing for a wider dissemination of our future studies.

Overall, we hope to disseminate our framework in the near-future with enough functionalities to support a wider array of research. Most the studies we conducted to date focused on discovering elements that negatively impacted the human-agent trust relationships. In future studies, we plan on testing different "trust repair" mechanisms to encourage users to re-engage with the agents after an initial negative experience. This focus call for the addition of new features to the agents, such as voice capabilities or for the agent to communication basic emotions.

Our publications [6, 7] show a growing interest in the development and assessment of human-agent relationships, with a focus on how users trust decision aids system. Participation in the Doctoral Consortium will allow us to get crucial feedback from the game research community on how to design for keeping players engaged with virtual agents in an interactive real-time task.

REFERENCES

- [1] Anthony L. Baker, Elizabeth K. Phillips, Daniel Ullman, and Joseph R. Keebler. 2018. Toward an understanding of trust repair in human-robot interaction: Current research and future directions. *ACM Transactions on Interactive Intelligent Systems* 8, 4 (2018), 1–30. <https://doi.org/10.1145/3181671>
- [2] Eric T. Chancey, James P. Bliss, Yusuke Yamani, and Holly A.H. Handley. 2017. Trust and the Compliance-Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence. *Human Factors* 59, 3 (2017), 333–345. <https://doi.org/10.1177/0018720816682648>
- [3] Alain Chavaillaz, Adrian Schwaninger, Stefan Michel, and Juergen Sauer. 2018. Automation in visual inspection tasks: X-ray luggage screening supported by a system of direct, indirect or adaptable cueing with low and high system reliability. *Ergonomics* (may 2018), 1–35. <https://doi.org/10.1080/00140139.2018.1481231>
- [4] Luca Chittaro and Fabio Buttussi. 2019. Exploring the use of arcade game elements for attitude change: Two studies in the aviation safety domain. *International Journal of Human-Computer Studies* 127 (2019), 112–123.
- [5] Filipa Correia, Carla Guerra, Samuel Mascarenhas, Francisco S Melo, and Ana Paiva. 2018. Exploring the Impact of Fault Justification in Human-Robot Trust. (2018), 507–513.
- [6] Sylvain Daronnat, Leif Azzopardi, and Martin Halvey. 2020. Impact of agents' errors on performance, reliance and trust in human-agent collaboration. <https://www.hfes2020.com/> Human Factors and Ergonomics Society Annual Meeting , HFES ; Accepted in press.
- [7] Sylvain Daronnat, Martin Halvey, Leif Azzopardi, and Mateusz Dubiel. 2019. Human-Agent Collaborations : Trust in Negotiating Control. *Workshop proceedings Everyday Automation Experience'19 In conjunction with CHI'19, May 5th, 2019, Glasgow, UK* (2019).
- [8] Mustafa Demir, Nathan J. McNeese, and Nancy J. Cooke. 2017. Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research* 46 (2017), 3–12. <https://doi.org/10.1016/j.cogsys.2016.11.003>
- [9] Katharina Emmerich, Patrizia Ring, and Maic Masuch. 2018. I'm Glad You Are on My Side: How to Design Compelling Game Companions. *The Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY '18* (2018), 141–152. <https://doi.org/10.1145/3242671.3242709>
- [10] Xiaocong Fan, Michael McNeese, Bingjun Sun, Timothy Hanratty, Laurel Allender, and John Yen. 2010. Human-Agent Collaboration for Time-Stressed Multicontext Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 40, 2 (mar 2010), 306–320. <https://doi.org/10.1109/TSMCA.2009.2035302>
- [11] Xiaocong Fan, Sooyoung Oh, Michael McNeese, John Yen, Haydee Cuevas, Laura Strater, and Mica R. Endsley. 2008. The influence of agent reliability on trust in human-agent collaboration. *Proceedings of the 15th European conference on Cognitive ergonomics the ergonomics of cool interaction - ECCE '08* (2008), 1. <https://doi.org/10.1145/1473018.1473028>
- [12] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. 2007. Measurement of trust in human-robot collaboration. *Proceedings of the 2007 International Symposium on Collaborative Technologies and Systems, CTS* (2007), 106–114. <https://doi.org/10.1109/CTS.2007.4621745>
- [13] Sandra G. Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society* (2006), 904–908. <https://doi.org/10.1177/154193120605000909>
- [14] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434. <https://doi.org/10.1177/0018720814547570>
- [15] Atari Inc. 1980. *Missile Command*. Game [Atari 2600]. Atari inc, Sunnyvale, USA. Last played July 2020.
- [16] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4, 1 (mar 2000), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04
- [17] Ariel Manzur Juan Linietsky. 2020. Godot Game Engine. <https://godotengine.org/>.
- [18] J. D. Lee and K. A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (2004), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- [19] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I Trust It, but I Don't Know Why. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 55, 3 (jun 2013), 520–534. <https://doi.org/10.1177/0018720812465081>
- [20] Richard Pak, Nicole Fink, Margaux Price, Brock Bass, and Lindsay Sturre. 2012. Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics* 55, 9 (2012), 1059–1072. <https://doi.org/10.1080/00140139.2012.691554>
- [21] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 2 (jun 1997), 230–253. <https://doi.org/10.1518/00187209778543886>
- [22] James Reason. 1990. Human error. *Cambridge University Press* (1990). <https://doi.org/10.7748/ns.11.49.20.s36>
- [23] J.K. Rempel, J.G. Holmes, and M.P. Zanna. 1985. Trust in close relationships scale. *Journal of Personality and Social Psychology* 49, 1 (1985), 95–112. <https://doi.org/10.1037/0022-3514.49.1.95>
- [24] Hirokazu Shirado and Nicholas A. Christakis. 2017. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* 545, 7654 (2017), 370–374. <https://doi.org/10.1038/nature22332>
- [25] Kimberly Stowers, Nicholas Kasdaglis, Michael Rupp, Jessie Chen, Daniel Barber, and Michael Barnes. 2017. *Insights into human-agent teaming: Intelligent agent transparency and uncertainty*. Vol. 499. 149–160 pages. https://doi.org/10.1007/978-3-319-41959-6_13