Integrating Humanoid Robots Into Simulation-Software-Generated Animations to Explore Judgments on Self-Driving Car Accidents*

Victoria Marcinkiewicz[a], Christopher D. Wallbridge[a], Qiyuan Zhang[a], Phillip Morgan[ab]
[a]Cardiff University Centre for AI, Robotics and Human-Machine Systems (IROHMS); Cardiff University Human Factors Excellence Research Group (HuFEx); School of Psychology, Cardiff University, 70 Park Place, Cardiff, CF10 3AT, UK; [b]Visiting Professor at Luleå University of Technology, Psychology, Division of Health, Medicine and Rehabilitation, Sweden.

***Abstract*— Building on the knowledge that human drivers (HD's) and self-driving cars (SDC's) are not blamed and trusted in the same way following a road traffic accident (RTA) or near-miss event, this paper proposes a novel method to investigate whether the manipulation of anthropomorphism in part using humanoid robots (HR) leads to reduced levels of blame and increased trust in SDC's that is more akin to HD's.**

I. Introduction

RTA's are largely attributable to human error. However, the emergence of SDC's could replace the need for humans to drive road vehicles (at least some of the time and/or under some conditions) consequently mitigating the source of many error-related RTA's caused by poor (human) driving.

The Society of Automotive Engineers defines 6 levels of driving automation: Level 0, no automation to Level 5 (L5), fully autonomous [1]. As the levels increase, so does the cars ability to drive itself under more conditions and circumstances eventually without the need for any human interaction at L5.

However, automation failure and RTA's remain a big concern even for L5 SDC's [2] which are still quite far off development and deployment. Despite promise from some that the technology will be superior to (most) human drivers, it will not be perfect and RTA's will still be inevitable.

Any adverse experience with an SDC (including RTA's, system failure, near-misses, etc.) is likely to erode human trust. Trust is a critical component in people's willingness to adopt new technologies and SDC's are no exception. A lack or indeed loss of trust will likely inhibit their uptake and adoption [3] and for some could lead to disuse [4]. It is therefore important to understand the factors which influence trust and blame assignment in an SDC following an incident.

Existing findings in this series of research have found that HD's and SDC's are not blamed and trusted in the same way following an incident or near-miss event – an SDC is usually blamed more than a human driver for executing the same actions under the same circumstances with the same consequences, compared to a human driver [5].

One explanation for this finding could be the tecnomorphic design of the SDC. With the dynamics of trust between people and robots only just beginning to be well-understood, Human Robot Interaction (HRI) research has largely suggested that increased anthropomorphism in a robot's design can promote trust [6]. Also, HR's can make SDC's appear more competent [7].

Despite this now well-established paradigm, a study by Onnasch et al (2022), suggests that an anthropomorphic robot design may not always universally promote trust in robots. Instead, it is argued that the successful implementation of anthropomorphic features is highly dependent on the context and the functionality fostered by the design [8].

With this in mind, the current paper proposes a novel method to investigate whether the level of anthropomorphism in an SDC causes it to be trusted/blamed in a way that is more akin to HD's. This will be achieved by integrating HR's into SDC's so they are perceived to be a part of the car. It may also be possible to determine whether an HR can be trusted more than a human driver.

Due to the nature of this research and the funding (UKRI ESRC- JST), there will also be the opportunity to build on existing work such as [9] to draw cross-country comparisons with Japan and in the future, other countries.

II. Proposed Methodology

A. Participants

The first study is being undertaken in the UK and Japan. A G-Power calculation [10] determined that at least 269 participants were needed from each country to detect a medium effect size (Cohen's f = 0.25) with power of 0.8.

Participation eligibility criteria include: aged ≥18-years; normal/normal-corrected vision and hearing; to speak English (UK data collection)/Japanese (Japan data collection) as a first language or be fluent as a second language. Participants will be recruited via Prolific, a globally trusted online recruitment platform.

B. Materials

Following Zhang et al's (2021) recognition that areas of SDC-accident research faces a huge methodological challenge - developing high-fidelity experimental stimuli as realistic representations of accident scenarios in order to elicit valid reactions from human participants [11] - this research will adopt the proposed 'Simulation-Software-Generated Animations' (SSGA) methodology.

*The work was funded as part of an ESRC-JST (Economic & Social Research Council - Japan Science & Technology Agency) project ES/T007079/1: Rule of Law in the Age of AI: Principles of Distributive Liability for Multi-Agent Societies. This work was also conducted with support of the Centre for Artificial Intelligence, Robotics and Human-Machine Systems (IROHMS) operation C82092 and partially funded by the European Regional Development Fund (ERDF) through the Welsh Government. Professor Phil Morgan (Cardiff University) is the UK Principal Investigator. Professor Tatsuhiko Inatani (Kyoto University) is the Principal Investigator for Japan with other collaborators from the Universities of Kyoto, Osaka and Doshisha.

Unlike research in other areas of HRI, attitudes like trust and acceptability of SDC's following an RTA cannot be measured after a real interaction. This is because it is not only impractical since L5 SDC's are still being developed but it is also ethically questionable whether participants should be explicitly exposed to such incidents.

One alternative method would be to use actual footage of an RTA but appropriate videos rarely exist; are hard to experimentally manipulate; do not always meet specific experiment objectives and can (e.g. in the event of accidents) be distressing for the participant. SSGA's however, strike a good balance between realism and practicality and overcome the above challenges therefore providing a suitable methodology for researching incidents involving SDC's [11].

For the current experiment, SSGA's also permit the novel integration of HR's to allow for the manipulation of anthropomorphism in the SDC. The SSGA's created depict a driving scenario of an SDC maneuvering around a bus. A HR robot was added to the bottom LHS corner of the screen, with an angle as if the HR were sitting on the dashboard. Participants were told that the HR was a part of the SDC's system. Each animation had varying levels of anthropomorphism (see Fig. 1.) but all culminated in the SDC hitting a pedestrian (note - the amination shows a freeze frame before this occurs with a description of the outcome). The study is currently taking place online.

C. Design

To operationalize anthropomorphism, the first experiment used a 3 (Conversation Style) x 2 (Presence of HR) between-subject design based on work by [7] (see Fig. 1). Our hypotheses were:

- H1 - As levels of speech increased (from no speech to
- conversational) trust in SDC's would increase.
- H2 - The presence of a HR would increase trust.
- H3 - There will be in interaction between presence of the HR and conversation style on blame.

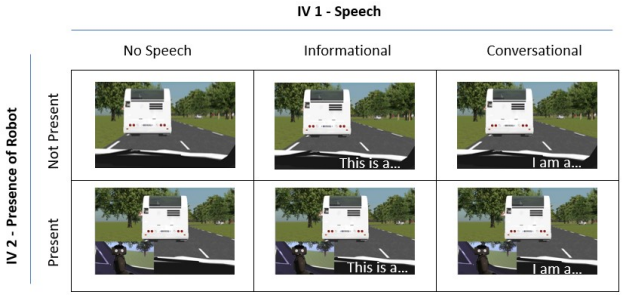

Figure 1. Example 2x3 Design Using Animation Methodology

D. Procedure

Participants were presented with an online information sheet explaining the aims; requirements; anonymizing of data process and their right to withdraw. Each participant was then asked if they wished to consent to partaking. Should the participant consent, they were asked to fill in a short preliminary questionnaire consisting of tick-box questions about their demographics including questions on gender; age; driving experience; how likely they are to use an SDC and to what extent they currently trust SDC technology.

Next, participants were required to watch one of six scenarios (randomly selected by the platform). They were then asked a series of self-report style questions based on the animation about how much they trusted the SDC after the RTA and who/what was to blame for what occurred.

Deciding on a suitable trust scale presented a second methodological challenge. For example, it has been recognized by Holthausen et al. (2020) that there is not a standardized method to measure trust in an SDC [12]. As a result the Situational Trust Scale for Autonomous Driving (STS-AD) [12] was the main scale used. Blame on both the AV and third parties was measured using questions based on [5]. Initial findings are currently undergoing analysis.

ACKNOWLEDGEMENT

This paper is dedicated to our dear friend and colleague Professor Dylan M Jones OBE DSc (30th Mar 1948 - 8th Apr 2022), without whom this work would not have been possible. Dylan had been at Cardiff University for almost 50-years leading Human Factors Psychology.

REFERENCES

[1] SAE International, "SAE Levels of Driving Automation Refined for Clarity and International Audience" (3-May-2021) [Online] Available: https://www.sae.org/blog/sae-j3016-update [Accessed: 9-June-2022]

[2] PA H.(2019). Some pitfalls in the promises of automated and autonomous vehicles, Ergonomics, 62(4), 479–495, doi: 10.1080/00140139.2018.1498136.

[3] Kim P.H, Dirks K.T, and Cooper, C.D. (2009). The repair of trust: A dynamic bilateral perspective & multilevel conceptualization, Acad. Manag. Rev, 34(3), 401–422, doi: 10.5465/AMR.2009.40631887.

[4] Parasuraman, R and Riley, V. Humans and automation: Use, misuse, disuse, abuse, Hum. Factors (1997)doi:10.1518/001872097778543886

[5] Zhang, Q., Wallbridge, C.D., Jones, D.M. and Morgan, P.L. (2021). The blame game: Double standards apply to autonomous vehicle accidents. In 12th International Conference on Applied Human Factors and Ergonomics, 308-314. Springer, Cham. Winner of the Best Paper award in the 9th International Conference on Human Factors in Transportation.

[6] Lee, J. G., Kim, K. J., Lee, S., & Shin, D. H. (2015). Can Autonomous Vehicles Be Safe and Trustworthy? Effects of Appearance and Autonomy of Unmanned Driving Systems. International Journal of Human-Computer Interaction, 31(10), 682–691.

[7] Lee, S.C., Sanghavi, H., Ko, S. and Jeon, M. (2019). Autonomous driving with an agent: Speech style and embodiment. In Proceedings of the 11th International Conference on Automotive User Interfaces Interactive Vehicular Applications: Adjunct Proceedings, 209-214.

[8] Onnasch, L., & Hildebrandt, C. L. (2022). Impact of Anthropomorphic Robot Design on Trust and Attention in Industrial Human-Robot Interaction. ACM Transactions on Human-Robot Interaction, 11(1) doi.org/10.1145/3472224

[9] Yun, Y., Oh, H. and Myung, R., 2021. Statistical Modeling of Cultural Differences in Adopting Autonomous Vehicles. Applied Sciences, 11(19), p.9030.

[10] Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

[11] Zhang, Q., Wallbridge, C.D., Morgan, P. and Jones, D.M. (2022). Using Simulation-software-generated Animations to Investigate. In 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022). In Print.

[12] Holthausen, B., Wintersberger, P., Walker. B. and Riener, A. (2020). Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation in Proceedings - 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI.