

Traveling Through the Dark: Using an interdisciplinary theatre and cognitive science approach to identify design strategies for human-machine shared experience in a self-driving car

Maurice Lamb¹, Rebecca Rouse¹

¹*Högskolan i Skövde, Institutionen för informationsteknologi*
Maurice.Lamb@his.se

*Traveling through the dark I found a deer
dead on the edge of the Wilson River road.
It is usually best to roll them into the canyon:
that road is narrow; to swerve might make more dead.*

-Excerpt from poem "Traveling through the Dark" by William Stafford (1998)

Recent research on human-machine interaction (HMI) across a range of fields, including both cognitive science and theatre, has stressed the need to re-frame such interactions as relational and based in shared experience (Gaggioli et al., 2021; Sciutti et al., 2018). In this case, the machine, whether software or hardware based, is characterized as an interaction partner instead of a tool. Reconceiving HMI as involving reciprocity and shared experiences moves away from transactional or one-sided models of interaction and requires exploring what can be meant by reciprocation and shared experience with a non-human partner. In particular, the concept of shared experience in HMI has been relatively under-explored due to both the typical framing of trust in HMI research and technological limitations of HMI systems. Refocusing the design of HMI systems on the ethos of shared experience can be supported by interdisciplinary research with theater. Anthropology and theater scholars have identified shared experience, termed 'communitas,' as a core component of the ontology of theater (Turner 1969). As such, theater has a long tradition of highly developed techniques for fostering shared experience. In addition, theater has a tradition of bringing the non-human 'to life,' as in puppetry and mask work (Bell 2001). Theater know-how about bringing humans and non-humans into meaningful relation via shared experience may be of use for the development in HMI design strategies, particularly when invited into interdisciplinary research collaboration early in the process, to allow for more meaningful exchange. We plan an exploratory pilot study where we use an interactive theatre approach to develop a set of dramatic scenarios, with the aim of identifying strategies for creating a sense of shared experience in human-machine relations, in the context of a self-driving car simulator. Specifically, using a Wizard of Oz setup we plan to develop scenarios aimed at providing varying degrees of shared experience with an autonomous vehicle. We will then have participants work with the faux autonomous vehicle in order to address a situation with an ethical component. We expect to gain clearer insights into how shared experiences shape trust in HMI and methodological limits/possibilities of the general approach. In the remainder of this abstract we will 1) introduce the recent discussion of reciprocity and shared experience in HMI, 2) discuss the intersection of theater and science, and 3) discuss the concept of morally permissible actions in the context of autonomous vehicle research.

Human-Machine Collaborators as Intersubjective Partners

Andrea Gaggioli et. al. (2021) make a compelling argument for reframing machine collaborators as partners instead of tools: "Effective collaboration in humans stems from 'growing together,' that is, from building a mutual understanding, which evolves over long periods through shared experiences. Consistently, we argue that the introduction of effective robot collaborators hinges upon the development of an intersubjective space between humans and machines" (Gaggioli 2021, 358). Gaggioli et. al. emphasize the machine (in their case, "robotic agent") as already a cultural artifact, acknowledging the creative intersubjective interplay between humans and objects always already in play. While it is acknowledged that humans often take an intentional stance towards non-human entities, the ways in which the intentional stance allows for the reflexive power of such entities to influence humans is often ignored. Following Gaggioli et. al., a dialogic experience structure may be most effective in fostering intersubjective relations

between humans and machines (Gaggioli 2021). Theater often features dialogic, narrative, and playful structures, and therefore a theatrical approach may be well-positioned to investigate intersubjective shared experience design for humans and machines.

Theater as Laboratory

Theater has an interesting history as a laboratory for innovation and invention. As a particularly plastic art, theater has the capacity to both incorporate and explore insights from scientific and technology disciplines in active ways, both through incorporating innovations in these fields to production techniques and by engaging core questions from these fields with audiences (Baugh 2014; Rouse 2018). The Bauhaus theater works from the early 1920s were used to explore larger issues about relationships between humans, technology, and affect while also innovating the theatrical form itself (Gropius and Wensinger 1961). Other well-known examples include the landmark “9 Evenings” collaboration between artists and Bell Labs engineers led by Billy Klüver and Robert Rauschenberg in 1966, and more recent work by the Blast Theory collective at University of Nottingham (Benford and Giannachi 2011). These types of art and science collaborations in the theatrical context have a long history which can be traced back at least to early medieval examples (Hoxby 2019, 161). Today, theater is often at the cutting edge of innovation in uses of new technologies, meaning theater today is often a fusion of the live and the digital (Dixon 2007). This positions theater as a medium uniquely suited to reflect, (re)enact, and share our own postdigital lived realities, in which our existence is marked by a heightened awareness of our always shifting entanglement of human, technological, and nonhuman in interconnected negotiation (Haraway 2016).

Autonomous Vehicles and Moral Permissibility

Discussions of human-machine relations often include questions of ethics, such as, how transparent should machine goals be to human interactors (Zonca & Sciutti 2021), and what is the role of encoded bias (Gitelman 2013, Eubanks 2018, Noble 2018, Hamraie 2017). Of particular interest in the current context are ethical issues relating to the kinds of actions a machine, the autonomous car in particular, will take in certain situations. Trolley problems typically take center stage in this kind of discussion, but in many of these types of scenarios the vehicle will not have time to interact with the person, and, on many accounts, the decision will be predetermined algorithmically. For this reason we aim to explore cases of morally permissible action, which are much more likely to occur and provide opportunity for reciprocal HMI. As in William Stafford’s poem above, a driver may be faced with several ethical dilemmas involving moral permissibility. Initially the driver in the poem is faced with at least 2 morally permissible options: 1) stop and move the animal, or 2) continue on. While many might prefer that the driver stop and move the animal, it is unclear if such action would be morally obligated. In many cases, it is at least permissible to pass the animal and carry on. Later in the poem, the driver is faced with the decision to save an animal or allow it to die, again a case of morally permissible options. These kinds of scenarios are far more likely to be opportunities for interaction between future drivers and autonomous vehicles. An even more common situation is choosing between different routes which may have a slightly bigger ecological impact, may result in shorter drive times, may affect the chances of congestion, and so on. In this kind of situation, we do not necessarily want to succumb to algorithmic morality and yet the decision has an inescapable ethical dimension (Parvin 2017, 317). A possible interactive solution could conceive of the vehicle as a kind of interactive expert system, offering suggestions and working through the reasons for and against a particular option. In this scenario, there will be a need for mutual trust, perhaps built on shared experiences traveling together. As an initial exploration of this traveling together, we envision an interactive performance which provides a space of shared imagining. In this case theatre can provide a platform for technologists, artists, and the public to come together in order to imagine possible futures for human-machine relations, and provide insights into an experience with possible intelligent HMI partners.

Conclusion: Experimental Performance Design

Using an immersive car simulator platform and Wizard of OZ (WoZ) methods, we plan to run a prototype interactive performance experience as a pilot study in identifying design strategies for shared experience in human-machine relations. We will construct a set of scenarios in which participants will have an opportunity to ride in an autonomous vehicle simulator and interact with a candidate AI, played by an actor. Participants may be with a more or less interactive agent with the aim of creating varying levels of shared experience. Towards the end of the ride the participant will face a scenario in which ethically valanced actions must be taken but which may involve several morally permissible choices. Exit interviews will allow for a clearer picture of participant experiences and guide development of a more rigorous scientific exploration. In line with research ethics guidelines, consent procedures will be followed and a thorough debriefing, including a clear discussion of the WoZ method employed will be provided (Rosén et al., 2021). In addition to providing initial research into design strategies for HMI shared experience, our project will also contribute knowledge toward methods for interdisciplinary art and science research collaborations.

References

- Baugh, C. (2014). *Theatre, Performance and Technology: the development and transformation of scenography*. Macmillan.
- Benford, S., Giannachi, G. (2011) *Performing Mixed Reality*. MIT Press.
- Dixon, S. (2007). *Digital Performance: a History of New Media in Theater, Dance, Performance Art, and Installation*. MIT Press.
- Eubanks, V. (2018) *Automating Inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Gaggioli, A., Chirico, A., Di Lernia, D., Maggioni, M. A., Malighetti, C., Manzi, F., Marchetti, A., Massaro, D., Rea, F., Rossignoli, D., Sandini, G., Villani, D., Wiederhold, B. K., Riva, G., & Sciutti, A. (2021). Machines Like Us and People Like You: Toward Human–Robot Shared Experience. *Cyberpsychology, Behavior, and Social Networking*, 24(5), 357–361. <https://doi.org/10.1089/cyber.2021.29216.aga>
- Gitelman, L. Ed. (2013) *Raw Data is an Oxymoron*. MIT Press.
- Hamraie, A. (2017). *Building Access: universal design and the politics of disability*. University of Minnesota Press.
- Haraway, D. (2016) *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press.
- Hoxby, B. (2019). Technologies of Performance: From Mystery Plays to the Italian Order. In: *A Cultural History of Theatre in the Early Modern Age*. Bloomsbury. 161-182.
- Mou, W., Ruocco, M., Zanatto, D., & Cangelosi, A. (2020). When Would You Trust a Robot? A Study on Trust and Theory of Mind in Human-Robot Interactions. *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 956–962. <https://doi.org/10.1109/RO-MAN47096.2020.9223551>
- Noble, S. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
- Parvin, N. (2017) Our Bodies in the Trolley's Path, or Why Self-Driving Cars Must Not be Programmed to Kill. *Science, Technology & Human Values* 43(2) 302-323.
- Rosén, J., Lindblom, J., Billing, E., & Lamb, M. (2021). *Ethical Challenges in the Human-Robot Interaction Field*. The Road to a successful HRI: AI, Trust and ethics - TRAITS Workshop, in conjunction with the 2021 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, USA, March 8--12 2021.
- Rouse, R. (2018) Partners: Human and Nonhuman Performers and Interactive Narrative in Postdigital Theater. *Interactive Digital Storytelling: LCNS 11318*. Springer. 369-382.
- Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine*, 37(1), 22–29. <https://doi.org/10.1109/MTS.2018.2795095>
- Turner, Victor. 1969. *The Ritual Process: Structure and Anti-Structure*. Chicago: Aldine Press.