International Conference “Trusting in Care Technology or Reliance on Socio-technical Constellation?”, February 15/16, 2024, Delmenhorst, Germany

Trusting robots & avatars A model of trust building dynamics with focus on embodied artificial agents

Philipp Graf[1,a], Manuela Marquardt[2]

[1] University of Applied Science Munich, Germany
[2] Charité – Universitätsmedizin Berlin, Germany
[a] Corresponding author; E-Mail: philipp.graf@hm.edu

Trust as a social mechanism to reduce complexity is a paramount factor in the context of care and medicine – especially when it comes to the use of technology for vulnerable patient groups. While personal trust is based on the expectation that another person, whom I trust, will arrive at the same assessment of behaviour and action alternatives given the same initial conditions of a situation (Luhmann 1968), trust in technology is based on the social construction that machines apparently "trivially" produce a stable output (Wagner 1994). Both the self-determined use of technology and trusting another person implies the risk of disappointment: the person or technology may not do what I expect. In the field of care and medicine both personal trust and trust in technology are called for, and they also take on a special relevance due to the strong physical involvement: The actions and behaviour of people and machines can have a negative impact on health, which increases the need for trust. Empirical evidence shows that the use of care technology therefore requires an active construction of trust in a situation in order to create a functioning and thereby safe interaction, for example to use a lifting device without unsettling people with dementia (Hornecker et al. 2020). Trust and wellbeing (and also safety) hereby are codependent.

The case of artificial agents, such as robots or virtual agents, adds another layer of complexity, as personal and technical aspects of trust are intertwined - the ontological status of these agents oscillates (Alač et al. 2011) and, according to our thesis, the emergence of personal trust and/or trust in technology experiences similar oscillations: It may or may not come to the emergence of mutual expectations (Lindemann 2009) as a precondition for personal trust, confidence in technical functioning or the violation of one of those or both. Empirical results show, for example, that artificial agents are not perceived as a coherent unit of body, mind, and identity, and that trust is thus divided into different loci of trust. Similarly, significant differences in the potential risks of trust can be identified depending on whether an agent is physically or virtually embodied (Mutlu 2021). Mutlu argues that different questions arise whether an artificial agent is encountered in a virtual or physical frame of mind.

The planned poster contribution will develop a model that focuses on the differences in the foundations and implications of trust-building dynamics between physically and virtually embodied agents deployed in the context of care or medicine. We expect differences depending on the stated purpose of an artificial agent deployment, where on the one hand it is about minimising risks and on the other hand it is about enabling the fulfillment of the purpose in a targeted way. We also expect differences in the spatio-temporal dimension: The fact that virtual agents draw the user into their virtual environment and at the same time cannot pose a physical threat has a positive effect on trust building when it comes to nonphysical purposes – such as data collection. Virtual agents can benefit from their artificial status because they are not seen as morally judgmental beings. On the downside, they rely heavily on temporal dynamics to handle interactions (Knorr Cetina 2009). Physically embodied agents, such as robots, on the other hand, invade the user's space. In this way, they are able to act co-presently and can affect their environment thus physically, but at the same time they pose an immense risk that requires the establishment of a basis of trust. In the context of care, trust-building dynamics are also embedded in and moderated by members of the organisation.

References

Alač, M. (2016). Social robots: things or agents? *AI & Society*, 31(4), 519–535. https://doi.org/10.1007/s00146-015-0631-6

Hornecker, E.; Bischof, A.; Graf, P.; Franzkowiak, L.; Krüger, N. (2020). The interactive enactment of care technologies and its implications for human-robot-interaction in care. In: David Lamas, et al. (eds), *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*.

Knorr Cetina, K. (2009). The synthetic situation: interactionism for a global world. *Symbolic Interaction*, 32(1), 61–87. https://doi.org/10.1525/si.2009.32.1.61

Lindemann, G., (2009): Die Verkörperung des Sozialen: Theoriekonstruktion und empirische Forschungsperspektiven, in: dies., Das Soziale von seinen Grenzen her denken. Weilerswist: Velbrück Wissenschaft, 162–181.

Luhmann, N. (1968). Vertrauen. Stuttgart: Enke.

Mutlu, B. (2021). The virtual and the physical: two frames of mind. *iScience*, 24(2), 101965. https://doi.org/10.1016/j.isci.2020.101965

Wagner, G. (1994). Vertrauen in Technik. *Zeitschrift für Soziologie*, 23(2), 145–157. https://doi.org/10.1515/zfsoz-1994-0205

Williams, T.; Ayers, D.; Kaufman, C.; Serrano, J.; Roy, S. (2021). Deconstructed trustee theory: disentangling trust in body and identity in multi-robot distributed systems. ACM/IEEE International Conference on Human-Robot Interaction, 262–271. https://doi.org/10.1145/3434073.3444644