

Blurring the line between human and robot? Mapping and manipulating the socialness gradient in the brain

ABSTRACT:

The process of understanding the minds of other people, such as their emotions and intentions, is mimicked when individuals try to understand an artificial mind. The assumption is that anthropomorphism, attributing human-like characteristics to non-human agents and objects, is an analogue to Theory-of-Mind, the ability to infer mental states of other people. Here, we test to what extent these two constructs formally overlap. Specifically, using a multi-method approach, we test if and how anthropomorphism is related to Theory-of-Mind using brain (Experiment 1) and behavioural (Experiment 2) measures. In a first exploratory experiment, we examine the relationship between dispositional anthropomorphism and activity within the Theory-of-Mind brain network ($n = 108$). Results from a Bayesian regression analysis showed no consistent relationship between dispositional anthropomorphism and activity in regions of the Theory-of-Mind network. In a follow-up, pre-registered experiment, we explored the relationship between Theory-of-Mind and situational and dispositional anthropomorphism in more depth. Participants ($n = 311$) watched a short movie while simultaneously completing situational anthropomorphism and Theory-of-Mind ratings, as well as measures of dispositional anthropomorphism and general Theory-of-Mind. Only situational anthropomorphism predicted the ability to understand and predict the behaviour of the film's characters. No relationship between situational or dispositional anthropomorphism and general Theory-of-Mind was observed. Together, these results suggest that while the constructs of anthropomorphism and Theory-of-Mind might overlap in certain situations, they remain separate and possibly unrelated at the personality level. These findings point to a possible dissociation between brain and behavioural measures when considering the relationship between Theory-of-Mind and anthropomorphism.

Keywords

Anthropomorphism, Theory-of-Mind, Social cognition, Attribution of socialness

Published Work:

de Jong, D., Hortensius, R., Hsieh, T.-Y., & Cross, E. S. (2021). Empathy and Schadenfreude in Human–Robot Teams. *Journal of Cognition*, 4(1), 35. doi: 10.5334/joc.177

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the age of human-robot interaction. *Trends in Neurosciences*, 43(6), 373-384. doi:10.1016/j.tins.2020.03.013

Hortensius, R., Kent, M., Darda, K. M., Jastrzab, L. E., Koldewyn, K., Ramsey, R., & Cross, E. S. (2021). Exploring the relationship between anthropomorphism and Theory-of-Mind in brain and behaviour. *Human Brain Mapping*, 42: 13. doi:10.1002/hbm.25542

Researcher's Contacts:

Dr. Ruud Hortensius

Email: ruud.hortensius@glasgow.ac.uk