Editorial: Do we really interact with artificial agents as if they are human?

Evelien Heyselaar[1]*, Nathan Caruana[2], Mincheol Shin[3], Leonhard Schilbach[4,5] and Emily S. Cross[2]

[1]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands, [2]Macquarie University, Sydney, NSW, Australia, [3]Department of Communication and Cognition, TSHD, Tilburg University, Tilburg, Netherlands, [4]Department of General Psychiatry 2, LVR-Klinikum Düsseldorf, Düsseldorf, Germany, [5]Ludwig Maximilian University of Munich, Munich, Bavaria, Germany

Editorial on the Research Topic
Do we really interact with artificial agents as if they are human?

Social interactions with artificial agents, such as voice agents, physically-embodied robots and avatars in virtual reality, are becoming increasingly normalised. As we strive to understand and optimise these social interactions–and human interactions in general–a pertinent question is: *Do we really interact with artificial agents as if they are human?* A wealth of related questions that are ripe for exploration concern the factors or conditions that might make this more or less likely.

In this Research Topic, we propose that this line of empirical enquiry is important, not only in informing how we can best design and position artificial agents in various applied contexts (e.g., education, entertainment, healthcare delivery), but also so we can inform how artificial agents can continue to be used as a valid tool in human social neuroscience research. Over the past decade, artificial agents have become a critical tool in experimental social neuroscience. In particular, virtual agent and virtual interaction paradigms have enabled social neuroscientists to achieve a balance between the need for 1) **ecological validity** on the one hand, with paradigms that capture the dynamic and reciprocal complexity of social interactions; and 2) **experimental control and objectivity**, with the ability to deploy paradigms in controlled laboratory and neuroimaging settings (that are typically designed to test one person at a time), with objective measures of social attention, behaviour and corresponding neural processes. Historically, studies of human social interaction have either used naturalistic and observational approaches that achieve 1) but not 2), or contrived and simplistic experimental studies–typically involving the passive observation of social information from a third person perspective–that achieve 2) but not 1). Recent calls for more interactive, second person neuroscience approaches have been met with the use of artificial agents and virtual interaction paradigms (Schilbach et al., 2013; Caruana et al., 2017c). Across this nascent body of research, it has largely been assumed that the neural, cognitive, and psychological mechanisms supporting social interactions between humans flexibly generalize to interactions with artificial agents and that they therefore can provide an ecologically-valid analogue for investigating these mechanisms. However, emerging research has highlighted that there are many factors, such as agent features (Cross and Ramsey, 2021; Henschel et al., 2021; Marchesi et al., 2021) or our beliefs and expectations about the agency and intentions of artificial agents (Klapper et al., 2014; Cross et al., 2016; Caruana et al., 2017a; Caruana et al., 2017b; Caruana and

McArthur, 2019), which can shape the extent to which the mechanisms of social cognition generalise across interactions with humans and artificial agents. We argue that this line of inquiry will synergistically inform best practice in the use of artificial agents as a tool for social interaction research, and in turn empirical insights will inform the conditions in which we can expect optimal and/or human-like interactions with artificial agents in our world.

The article by Kyrlitsias and Michael-Grigoriou (this volume) gives an in-depth overview of the various moderators that play a role in the perception of social interaction with artificial agents in a virtual environment. Notably, the authors mention how variables such as human-like movement and technological advancements can influence the feeling of "presence" and "co-presence." That virtual agents with more human-like social ability make the user feel more comfortable. But what is the ultimate goal? What level of co-presence is similar to what we experience when we interact with other humans?

The article by Huang and Moore (this volume) gives their opinion on these Research Topic directly. Their introduction even states: "It is hoped in many studies that robots designed with anthropomorphic appearances and human-like cognitive behaviours can enable humans to interact with them in similar ways as they would interact with other humans, even to develop social bonds." The article then goes on to describe how human-human interaction is not formulaic, like that of an artificial agent. A social robot's human-like affordances could therefore be seen as "dishonest" because it hides the fact that a virtual agent is a "mismatched" conversational partner. The authors, therefore pose an additional, yet related, empirical question: Does a virtual partner need to be human-like in order to improve human users' experience in HRI? Highlighting the need for research in this space to consider the contexts in which perceiving artificial agents as human or human-like is likely to be necessary or optimal.

Whereas Huang and Moore focus mainly on robots, Huang and Jung (this volume) extend this concept of authenticity to all virtual characters. This is a timely proposition, considering the rapid increase in commercially available artificial intelligence applications, such as ChatGPT, Vocaloid virtual idols and other emerging virtual characters that blur the boundary between human and machine agency as well as human "authenticity." The current paper focusses on the important, but often overlooked, element of authenticity and highlights the need for a unified theoretical framework to guide empirical research in this space.

Beyond providing successful and productive interactions with digital agents, social virtual entities are also used in research to inform human-human interaction. Given the large amount of experimental control and logistical ease that using a digital confederate provides, it is no surprise that it has become a common experimental methodology. In these instances, it is of utmost importance that participants interact with the digital agent as if it were another human. Gregory et al. (this volume) provides arguments

from the literature showing that there are many tasks in which participants show little behavioral difference whether they are doing the task with a human or with a virtual agent. This work again shows the likely highly context-specific nature of the role of these beliefs and expectations in shaping interactions with artificial agents.

A second goal of this Research Topic was to determine whether the underlying neurocognitive mechanisms that are engaged during human-agent social interactions are the same (or comparable) to those in human-human interactions. Suggestions brought forth by the articles described above indicate that perhaps the variation that exists across humans, across artificial agents and across the various contexts in which humans might interact with human/artificial agents, makes this simple comparison reductionistic. This line of enquiry will require a careful appreciation of this variability, especially given the gradation that exists in how we conceptualize artificial agents (see Cross and Ramsey, 2021 for related discussion). For instance, a simple artificial agent that is predictably programmed *versus* an artificial agent that is self-learning and adapts its behavior over time are likely to load on very different social-cognitive mechanisms when engaged in a social interaction, and result in markedly different social outcomes. As such, how we define artificial agents, and categorize them by their features and context for application, is likely to be key in structuring this important line of enquiry.

Author contributions

NC and EH made the first draft of the editorial, the other authors gave useful revisions. All authors read and approved the editorial before submission.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Caruana, N., de Lissa, P., and McArthur, G. (2017a). Beliefs about human agency influence the neural processing of gaze during joint attention. *Soc. Neurosci.* 12 (2), 194–206. doi:10.1080/17470919.2016.1160953

Caruana, N., and McArthur, G. (2019). The mind minds minds: The effect of intentional stance on the neural encoding of joint attention. *Cognitive, Affect. Behav. Neurosci.* 19 (6), 1479–1491. doi:10.3758/s13415-019-00734-y

Caruana, N., McArthur, G., Woolgar, A., and Brock, J. (2017b). Simulating social interactions for the experimental investigation of joint attention. *Neurosci. Biobehav. Rev.* 74, 115–125. doi:10.1016/j.neubiorev.2016.12.022

Caruana, N., Spirou, D., and Brock, J. (2017c). Human agency beliefs influence behaviour during virtual social interactions. *PeerJ* 5, e3819. doi:10.7717/peerj. 3819

Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., and Hamilton, A. F. de C. (2016). The shaping of social perception by stimulus and knowledge cues to human animacy. *Philosophical Trans. R. Soc. B Biol. Sci.* 371 (1686), 20150075. doi:10.1098/rstb.2015.0075

Cross, E. S., and Ramsey, R. (2021). Mind meets machine: Towards a cognitive science of human–machine interactions. *Trends Cognitive Sci.* 25 (3), 200–212. doi:10.1016/j.tics.2020.11.009

Henschel, A., Laban, G., and Cross, E. S. (2021). What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Curr. Robot. Rep.* 2 (1), 9–19. doi:10.1007/s43154-020-00035-0

Klapper, A., Ramsey, R., Wigboldus, D., and Cross, E. S. (2014). The control of automatic imitation based on bottom–up and top–down cues to animacy: Insights from brain and behavior. *J. Cognitive Neurosci.* 26 (11), 2503–2513. doi:10.1162/jocn_a_00651

Marchesi, S., Bossi, F., Ghiglino, D., De Tommaso, D., and Wykowska, A. (2021). I Am looking for your mind: Pupil dilation predicts individual differences in sensitivity to hints of human-likeness in robot behavior. *Front. Robotics AI* 8, 653537. doi:10.3389/frobt.2021.653537

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36 (4), 393–414. doi:10.1017/S0140525X12000660