

 AccessEnglish | **Français**

Autonomous social robots are real in the mind's eye of many

Published online by Cambridge University Press: **05 April 2023**In response to: **Social robots as depictions of social agents**[Related commentaries](#) (29) [Author response](#)Nathan Caruana  and Emily S. Cross 

Nathan Caruana

Affiliation: School of Psychological Science, Macquarie University, Sydney, NSW, Australia
nathan.caruana@mq.edu.au <https://nathancaruaana.weebly.com/>

Emily S. Cross

Affiliation: School of Psychological Science, Macquarie University, Sydney, NSW, Australia
nathan.caruana@mq.edu.au <https://nathancaruaana.weebly.com/>; Centre for Elite Performance, Expertise and Training, Macquarie University, Sydney, NSW, Australia
emily.cross@mq.edu.au <https://www.soba-lab.com/>; Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK; MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Westmead, NSW, Australia[Article contents](#)

Abstract

Clark and Fischer's dismissal of extant human–robot interaction research approaches limits opportunities to understand major variables shaping people's engagement with social robots. Instead, this endeavour categorically requires multidisciplinary approaches. We refute the assumption that people cannot (correctly or incorrectly) represent robots as autonomous social agents. This contradicts available empirical evidence, and will become increasingly tenuous as robot automation improves.

Type

Open Peer Commentary

InformationBehavioral and Brain Sciences, Volume 46, 2023, e26DOI: <https://doi.org/10.1017/S0140525X22001625>**Copyright**

Copyright © The Author(s), 2023. Published by Cambridge University Press

Clark and Fischer (C&F) claim that people represent social robots as *depictions* of other humans, and not as autonomous social entities. We argue this framework for understanding human perceptions of – and interactions with – robots is limited and limiting. Instead, an eclectic approach drawing upon psychology, social neuroscience, and human–robot interaction (HRI) will best serve empirical progress as robots' social capabilities evolve.

We agree that for some people, and in particular contexts, certain robots can be seen as representing the intentions and actions of a human principal (e.g., operator/engineer). Our central argument, however, is that such a framework for understanding HRIs is not universal and may become irrelevant as increasingly intelligent and autonomous social robots are realised.

In serving their claim, the authors draw upon Wizard-of-Oz approaches commonly used in HRI research (where a person teleoperates a robot) to categorise robots alongside ventriloquist dummies as examples of “interactive” depictions, which are a step above “staged” (e.g., puppets) and “static” (e.g., statues) depictions within their taxonomy. However, the fact that a robot “depicts” human intentions/actions *in reality* does not mean people perceive it as such. An overlooked feature of the Wizard-of-Oz approach is the use of Turing deception, in which people believe the robot operates autonomously (Kelley, 1984). We argue that under many circumstances, humans *do* perceive social robots as autonomous, intentional agents, even when they are not. In fact, several studies have demonstrated direct consequences on subjective experiences, behaviour and neural processing during interactions with virtual agents and robots depending on whether or not people believe an agent is human-controlled (Caruana & McArthur, 2019; Caruana, Spirou, & Brock, 2017; Cross, Ramsey, Liepelt, Prinz, & de C Hamilton, 2016; Schellen & Wykowska, 2019). These studies (and others) show that under some conditions, people represent artificial agents as human depictions, and sometimes not. Thus, it remains unclear how the

depictions framework can resolve such findings if human depictions are always in play. If the “variety” or “proximity” of the human depiction shapes experiences with robots, this requires further specification.

This also highlights the important roles played by knowledge and beliefs in shaping robot representations; that is, what a person *believes* a robot can do, which is often quite separate from what a robot *can actually do*. The authors touch on the issue of pretense, and correctly state that what children know and believe about robots is unclear. We would go further to state that this holds for people *of all ages*, highlighting another HRI research challenge unresolved by the depictions framework. Specifically, it is reasonable to assume that most people have clearly defined, relatively invariant, top-down knowledge cues concerning puppets' or ventriloquist dummies' autonomy and sentience. The presence of strings and/or the close proximity of the ventriloquist offer bottom-up cues that activate knowledge of these agents being directly operated by their human “principal.” The same, however, cannot be said of social robots, whose relationship with their principal(s) can be far more distant, ambiguous, opaque, or complex – especially upon first encounter. Social robots are also more novel and varied than puppets or ventriloquist dummies, and continue to evolve as technologies develop, further fuelling this ambiguity. The depictions framework does not accommodate for this ambiguity in robot agency, nor the variability in the kinds of cues humans rely on to resolve it. Further, many individuals are naive about the current state of robot capabilities, or biased by representations of autonomous social robots in popular media (cf. Cross & Ramsey, 2021). Indeed, our own recent research encourages the hypothesis that some children may have rather realistic ideas about the autonomy and limitations of robots (Caruana, Moffat, Blanco, & Cross, 2022). Furthermore, many contexts exist for applying socially assistive robots (e.g., education, health, and aged care) where users may be more likely to overestimate robots' autonomy, and perhaps less likely to see them as depictions of humans (e.g., young children, the elderly, and individuals with intellectual disabilities).

Rather than focusing on questions of depiction and how clearly or accurately people associate a robot with its human engineer(s), we argue that thornier challenges arise from issues related to variability in HRI across (1) *individual differences* (e.g., personality, knowledge, attribution styles, education, cognitive ability); (2) *robot form* (e.g., zoomorphic, mechanoid, humanoid, size, composition), and *function* (e.g., verbal, mobile, expressive); and (3) *application domains*. Together, this considerable variation and complexity presents deep challenges to building a robust knowledge base related to social encounters between humans and robots (Cross & Ramsey, 2021), and it remains unclear how this is resolved within the depictions framework. We argue that this problem requires a multidisciplinary,

eclectic approach receptive to insights gained from the previous approaches that C&F summarise and dismiss.

We would further argue that these approaches have not been fully represented in this review, and that they continue to bear fruit in explaining variability across HRIs. For instance, the review of the “trait attribution” approach loosely references Epley, Waytz, Akalis, and Cacioppo's (2008) concepts of “elicited knowledge” and “effectance motivation” for explaining why humans may ascribe human-like agency/intentions to objects. However, key to this approach is the idea that significant *individual differences* exist in people's tendencies to anthropomorphise, which these factors attempt to explain (e.g., Neave, Jackson, Saxton, & Hönekopp, 2015). Another overlooked component of Epley et al.' framework concerns “sociality motivation.” This refers to one's drive to be socially connected to others, and is argued to interact with the abovementioned factors to influence anthropomorphism, while also being influenced by other contextual or dispositional factors (e.g., subjective loneliness, social isolation, anxiety, personality, culture, etc.). While we fully acknowledge that none of the extant approaches for understanding HRI offers complete explanations for the “social artifact” problem, they nonetheless offer useful frameworks for understanding how some variables shape people's interactions with robots. They also continue to inspire new empirical questions that advance our knowledge of the factors that shape HRIs. To us, it remains unclear how the depictions framework offers a solution to the inadequacies of extant approaches, or hypotheses that will help advance the field.

Financial support


Dr. Caruana was supported by a Macquarie University Research Fellowship (9201701145). Professor Cross was supported in part by funding from the European Research Council (ERC) under the EU Horizon 2020 research and innovation programme (grant agreement 677270) and the Leverhulme Trust (PLP-2018-152). The funding sources for this project were not involved in any decision-making, data collection, analysis, or dissemination of this study. The scientific process of this study was independent of any relationship with the funding sources that could be construed as a potential conflict of interest.

Competing interest


None.

References


- ^ Caruana, N., & McArthur, G. (2019). The mind minds minds: The effect of intentional stance on the neural encoding of joint attention. *Cognitive, Affective & Behavioral Neuroscience*, 19(6), 1479–1491. <https://doi.org/10.3758/s13415-019-00734-y> [CrossRef](#) [Google Scholar](#) [PubMed](#) [Find it@MQ](#) [Find it@MQ](#)
- ^ Caruana, N., Moffat, R., Blanco, A. M., & Cross, E. S. (2022). Perceptions of intelligence & sentience shape children's interactions with robot reading companions: A mixed methods study. *PsyArXiv*. <https://doi.org/10.31234/osf.io/7t2w9> [Google Scholar](#) [Find it@MQ](#) [Find it@MQ](#)
- ^ Caruana, N., Spirou, D., & Brock, J. (2017). Human agency beliefs influence behaviour during virtual social interactions. *PeerJ*, 5, e3819. <https://doi.org/10.7717/peerj.3819> [CrossRef](#) [Google Scholar](#) [PubMed](#) [Find it@MQ](#) [Find it@MQ](#)
- ^ Cross, E. S., & Ramsey, R. (2021). Mind meets machine: Towards a cognitive science of human–machine interactions. *Trends in Cognitive Sciences*, 25(3), 200–212. [10.1016/j.tics.2020.11.009](https://doi.org/10.1016/j.tics.2020.11.009) [CrossRef](#) [Google Scholar](#) [PubMed](#) [Find it@MQ](#) [Find it@MQ](#)
- ^ Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., & de C Hamilton, A. F. (2016). The shaping of social perception by stimulus and knowledge cues to human animacy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1686), 20150075. <https://doi.org/10.1098/rstb.2015.0075> [CrossRef](#) [Google Scholar](#) [PubMed](#) [Find it@MQ](#) [Find it@MQ](#)
- ^ Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, 26(2), 143–155. [doi:10.1521/soco.2008.26.2.143](https://doi.org/10.1521/soco.2008.26.2.143) [CrossRef](#) [Google Scholar](#) [Find it@MQ](#) [Find it@MQ](#)
- ^ Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on*

Information Systems, 2(1), 26–
41.10.1145/357417.357420 [CrossRef](#) [Google Scholar](#)  [Find it@MQ](#)



Neave, N., Jackson, R., Saxton, T., & Hönekopp, J. (2015). The influence of anthropomorphic tendencies on human hoarding behaviours. *Personality and Individual Differences*, 72, 214–219.
<https://doi.org/10.1016/j.paid.2014.08.041> [CrossRef](#) [Google Scholar](#)  [Find it@MQ](#)



Schellen, E., & Wykowska, A. (2019). Intentional mindset toward robots – Open questions and methodological challenges. *Frontiers in Robotics and AI*, 5, 139.
<https://doi.org/10.3389/frobt.2018.00139> [CrossRef](#) [Google Scholar](#) [PubMed](#)  [Find it@MQ](#)

Related content

AI-generated results: by UNSILO

Chapter

Embodiment and Algorithms for Human–Robot Interaction

[The Cambridge Handbook of the Law of Algorithms](#)

Published online: 19 October 2020

Article

Systematic Generation of a 3D DSM by Extracting Social Robot Behaviors from Literature

Ilayda Ozer and Zuhail Erden

[Proceedings of the Design Society: International Conference on Engineering Design](#)

Published online: 26 July 2019

Chapter

Social Signal Processing in Social Robotics

Maha Salem and Kerstin Dautenhahn

[Social Signal Processing](#)

Published online: 13 July 2017

Chapter

Body Movements Generation for Virtual Characters and Social Robots

Aryel Beck, Zerrin Yumak and Nadia Magnenat-Thalmann

[Social Signal Processing](#)

Published online: 13 July 2017

Chapter

NOC NOC, Who's There? A New Ontological Category (NOC) for Social Robots

Peter H. Kahn, and Solace Shen

[New Perspectives on Human Development](#)

Published online: 11 May 2017

Article

I am a warm robot: the effects of temperature in physical human–robot interaction

Eunil Park and Jaeryoung Lee

[Robotica](#)

Published online: 2 August 2013

Chapter

What Is Human–Robot Interaction?

Christoph Bartneck, Tony Belpaeme, Friederike Eyssele, Takayuki Kanda, Merel Keijsers and Selma Šabanović

[Human-Robot Interaction](#)

Published online: 20 February 2020
