The second-order problem of other minds

Ori Friedman, University of Waterloo

Arber Tasimi, Emory University

This is an open access version of:

Friedman, O. & Tasimi, A. (2023). The second-order problem of other minds. *Behavioral and Brain Sciences*, *46*, E31. https://doi.org/10.1017/S0140525X22001443

Abstract.
The target article proposes that people perceive social robots as depictions rather than as genuine social agents. We suggest that people might instead view social robots as social agents, albeit agents with more restricted capacities and moral rights than humans. We discuss why social robots, unlike other kinds of depictions, present a special challenge for testing the depiction hypothesis.

Main text.
How will we know when a social robot (or any other kind of artificial intelligence) is a genuine social agent? That is, how will we know when it is conscious, feels things, and understands what it hears or says? This is the philosophical problem of other minds—the problem of how we can know that anyone else has a mind—applied to human creations (Harnad, 1991).

The target article raises a new problem of other minds. Clark and Fischer suggest that rather than viewing social robots as genuine social agents, people instead view them as depictions of social agents. Under this depiction account, people engage in a kind of pretense when interacting with social robots (also see Rueben et al, 2020). This account could be right, but we suggest it remains possible that people might instead view social robots as genuine social agents. Testing between these accounts introduces a new second-order problem of other minds: How can we tell if other people think they are dealing with a genuine social agent or a mere depiction of one?

The second-order problem of other minds may be difficult to resolve. When dealing with depictions, people normally hold back—their actions fall short from what they would do with the real thing. For example, children pretending to eat plastic fruit refrain from actually biting it (e.g., Leslie & Happé, 1989; Lillard, 1993) and filmgoers do not attempt to intervene in movie events. Do people also hold back with social robots? It might be hard to tell. Although people do not treat social robots exactly the way they treat their peers, this isn't saying much. There are many different kinds of agents and people see them as varying in their mental capacities (Grey et al., 2007; Weisman et al., 2017) and moral standing (Crimston et al., 2018; Goodwin, 2015). So, while it might be obvious when people hold back when dealing with many kinds of depictions (e.g., plastic fruit), this will be less obvious with social robots. What looks like holding back could turn out to reflect beliefs that social robots have limited capacities and moral standing.

To illustrate these points, let's consider the evidence offered as support for the idea that people view social robots as depictions. One line of evidence is that rather than seeing social robots the way they see their fellow humans, people see social robots as a kind of property. They affirm social robots can be sold, and if a social robot dented someone's car, the owner of the car would seek compensation from the robot's owner rather than from the robot itself. Treating social robots like property might follow from the belief that they are depictions rather than genuine social agents. But it is also reminiscent of how people treat real agents viewed as having limited moral standing or limited mental abilities. For example, pets and other animals are bought and sold and their owners are held liable when they cause harm (e.g., Bowman-Smith et al., 2018; Nadler & McDonnell, 2011). Similar points may apply to how enslaved people were viewed and

treated in the past. They too were treated as chattel, and when they caused harm, their enslavers were held liable in some legal systems (Oppenheim, 1940). But it is unlikely that people view pets as depictions, or that the enslaved were viewed this way either. So rather than viewing robots as mere depictions, people might instead see them as genuine agents with limited moral worth and limited mental capacities.

The target article also notes that people differ from one another in how they interact with social robots. Although some people converse with social robots, others refrain from doing so—these people do not respond to greetings from social robots and if they address robots at all, it is only with blunt questions and brusque orders; perhaps these people are unwilling to play along with the pretense that these depictions are social agents. But this again is ambiguous. We might also expect differences between people if some believed that social robots are real agents, while others did not. Here again, people's treatment of animals raises questions. As with social robots, people vary in how they address their pets and some people's communication with their pet dogs is apparently limited to commands and threats (e.g., Carlisle-Frank et al., 2004; Mitchell, 2004). Some talk to pets could have a pretend element—people sometimes ask dogs questions but then also answer the questions (Mitchell, 2004). But it seems unlikely (at least to us) that people view pets as depictions, or that most variation in talk to pets come down to differences in owners' proclivity to pretend.

Although the second-order problem of other minds be difficult to resolve, the difficulty may be asymmetric. While it might be difficult to confirm that social robots are viewed as depictions, it may be easier to confirm when they are viewed as genuine agents. Consider the issue of whether people show moral concern for robots (for a recent review see Harris & Anthis, 2021). When people express concerns for the welfare of robots and advocate for robots to have rights, this might suggest they view social robots as genuine agents – at least if these expressions of concern focus on robots themselves and not on side-concerns, such as concerns about property damage, or concerns that mistreating robots will encourage mistreatment of humans (e.g., Levy, 2009). By contrast, absence of concern would not necessarily show that people view robots as depictions. It could instead stem from viewing robots as genuine agents with limited capacities or moral worth.

Conflicts of Interest. None

Funding Statement. This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada awarded to OF.

Reference list.

Bowman-Smith, C. K., Goulding, B. W., & Friedman, O. (2018). Children hold owners responsible when property causes harm. *Journal of Experimental Psychology: General*, *147*(8), 1191-1199. https://doi.org/10.1037/xge0000429

Carlisle-Frank, P., Frank, J. M., & Nielsen, L. (2004). Selective battering of the family pet. *Anthrozoös*, *17*(1), 26-42. https://doi.org/10.2752/089279304786991864

Crimston, C. R., Hornsey, M. J., Bain, P. G., & Bastian, B. (2018). Toward a psychology of moral expansiveness. *Current Directions in Psychological Science*, *27*(1), 14-19. https://doi.org/10.1177/0963721417730888

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619-619. https://doi.org/10.1126/science.1134475

Goodwin, G. P. (2015). Experimental approaches to moral standing. *Philosophy Compass*, *10*(12), 914-926. https://doi.org/10.1111/phc3.12266

Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, *1*(1), 43-54. https://doi.org/10.1007/BF00360578

Harris, J., & Anthis, J. R. (2021). The moral consideration of artificial entities: a literature review. *Science and Engineering Ethics*, *27*(4), 1-95. https://doi.org/10.1007/s11948-021-00331-8

Leslie, A. M., & Happé, F. (1989). Autism and ostensive communication: The relevance of metarepresentation. *Development and Psychopathology*, *1*(3), 205-212. https://doi.org/10.1017/S0954579400000407

Levy, D. (2009). The ethical treatment of artificially conscious robots. *International Journal of Social Robotics*, *1*(3), 209-216. https://doi.org/10.1007/s12369-009-0022-6

Lillard, A. S. (1993). Pretend play skills and the child's theory of mind. *Child Development*, *64*(2), 348-371. https://doi.org/10.1111/j.1467-8624.1993.tb02914.x

Mitchell, R. W. (2004). Controlling the dog, pretending to have a conversation, or just being friendly? Influences of sex and familiarity on Americans' talk to dogs during play. *Interaction Studies*, *5*(1), 99-129. https://doi.org/10.1075/is.5.1.06mit

Nadler, J., & McDonnell, M. H. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review*, *97*, 255-304.

Oppenheim, L. (1940). Law of Slaves—A Comparative Study of the Roman and Louisiana Systems. *Tulane Law Review*, *14*, 384-406.

Rueben, M., Rothberg, E., & Matarić, M. J. (2020). Applying the theory of make-believe to human-robot interaction. In M. Nørskov, J. Seibt, & O. S. Quick (Eds.), *Culturally Sustainable Social Robotics* (pp. 40-50). IOS Press. https://doi.org/10.3233/FAIA200899

Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences*, *114*(43), 11374-11379. https://doi.org/10.1073/pnas.170434711