

Please cite as: Veit, W. & Browning, H. (2022). Social Robots and the Intentional Stance. Preprint.

Check www.walterveit.com for citation details once published

Social Robots and the Intentional Stance

Walter Veit & Heather Browning

Abstract:

Why is it that people simultaneously treat social robots as mere designed artifacts, yet show willingness to interact with them as if they were real agents? Here, we argue that Dennett's distinction between the intentional stance and the design stance can help us to resolve this puzzle, allowing us to further our understanding of social robots as interactive depictions.

Main text:

Clark and Fischer (2022) offer an excellent analysis of what they call the *social artifact puzzle*, i.e. why it is that people simultaneously (1) hold the view that social robots - whether in the shape of animals or humans - are merely designed mechanical artifacts, and (2) show willingness to interact with them as if they were real agents. Their solution to this apparent inconsistency is to suggest that people do not inherently treat social robots as real agents, but rather treat them as interactive depictions (i.e. analogues) to real agents. To our surprise, however, in their discussion the authors did not mention Daniel Dennett's (1987, 1988) distinction between the intentional stance and the design stance – two attitudes that humans routinely take in their engagement with the world. Yet, we think that it is precisely this distinction that can help to address some of the unresolved issues the authors raise as currently lacking from the alternative perspectives: why (i) people differ in their *willingness* to interact with social robots, (ii) why people can *rapidly change their perspective* of social robots, from agents to artifacts, and (iii) why people seem to only *selectively* treat social robots as agents.

The intentional stance, according to Dennett, involves treating “the system whose behavior is to be predicted as a rational agent; one attributes to the system the beliefs and desires it ought to have, given its place in the world and its purpose, and then predicts that it will act to further its goals in the light of its beliefs” (Dennett 1988, p. 496). This stance can be applied to other agents as well as to oneself (Veit et al. 2019, Veit 2022). On the other hand, when one takes the *design stance* “one predicts the behavior of a system by assuming that it has a certain design (is composed of elements with functions) and that it will behave as it is designed to behave under various circumstances” (Dennett 1988, p. 496).

When humans are faced with a social robot, both stances are useful for predicting how the robot is going to behave, so people are faced with a choice of how to treat it. Which stance they choose to adopt may depend on a range of factors, including individual differences, and the particular goals of

the interaction. For instance, people will differ in their social personality traits, and their prior experience with social robots or similar artificial agents, which makes it unsurprising that they will then also differ in their willingness to adopt the intentional stance and interact with them as if they were real agents with beliefs and desires; as opposed to adopting the design stance and treating them in a more pragmatic manner, as useful objects but nothing more (though we note that Marchesi et al. (2019) did not find any differences within the demographic groups they screened for).

Thinking about these perspectives as conditional and changing stances, rather than strong ontological and normative commitments about the status of social robots and how they should be treated, removes the mystery regarding why and how people can rapidly change their perspectives of social robots, treating them as artifacts at one point in time and as agents at another. It can now be regarded as a fairly simple switch from one stance to another. This also provides a solution to the question of why people show selectivity in their interpretation of the capacities and abilities of social robots. People can adopt one stance or the other, depending on the context and goals of the particular interaction.

It is important to keep in mind that both stances are ultimately meant to be useful within different contexts. Our interactions with social robots will occur within a range of contexts, and people will have vastly different goals depending both on their own aims and values, and the situation they are encountered in. In some cases it will be useful for someone, with reference to their goals, to ignore the non-human-like features of a social robot and treat them as another social agent. Particularly, in light of the evidence the authors discuss, of people's strong emotional responses to some social robots (e.g. companion 'animals'), there may here be psychological and social benefits in adopting the intentional stance and treating the robot as a social agent (indeed, this would appear to be the very purpose of these robots in the first place). It may also assist in rapid and flexible predictions of behaviour, supported by the fact that people more readily adopt the intentional stance when viewing social robots interacting with other humans, than when viewing them acting alone (Spatola et al. 2021). In other cases, often even within the same interaction, it will be more useful to ignore the human-like features and focus on the more mechanical properties, shifting to a treatment of the robot as an artifact instead. This is more likely in cases where interaction with the robot is more instrumental, in service of some other goal.

We want to emphasise that one doesn't have to see Dennett's account as a competitor to Clark and Fischer's. Indeed, we think they are complementary. Our suggestion here is that the authors could include this distinction within their proposal, drawing more links between their account and some of the existing studies that explore the intentional and design stances in relation to people's responses to robots (e.g. Marchesi et al. 2019; Perez-Osorio et al. 2019; Spatola et al. 2021). In particular, we see benefit in more empirical research on people's interactions with and attitudes towards social robots, to test these ideas and see which may apply more strongly within different contexts. As the current evidence base is small, and underdetermines the current available theories, if we want to advance our understanding of when, how and why ordinary people treat social robots as agents, we will ultimately need further empirical work and we think that Dennett's distinction provides an additional useful framework from which to build this.

Conflict of interest statement:

The authors have no conflicts of interest to report.

Funding Information:

WV's research was supported under Australian Research Council's Discovery Projects funding scheme (project number FL170100160).

References

Clark, H., & Fischer, K. (2022). Social robots as depictions of social agents. *Behavioral and Brain Sciences*, 1-33.

Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.

Dennett, D. C. (1988). Précis of the intentional stance. *Behavioral and brain sciences*, 11(3), 495-505.

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots?. *Frontiers in psychology*, 10, 450.

Perez-Osorio, J., & Wykowska, A. (2019). Adopting the intentional stance towards humanoid robots. In *Wording robotics* (pp. 119-136). Springer, Cham.

Spatola, N., Marchesi, S. & Wykowska, A. (2021). The intentional stance test-2: How to measure the tendency to adopt intentional stance towards robots. *Frontiers in Robotics and AI*. 8: 666586

Veit, W. (2022). Revisiting the Intentionality All-Stars. *Review of Analytic Philosophy*, 2(1), 1-24. <https://doi.org/10.18494/SAM.RAP.2022.0009>

Veit, W., Dewhurst, J., Dolega, K., Jones, M., Stanley, S., Frankish, K. & Dennett, D.C. (2019). The Rationale of Rationalization. *Behavioral and Brain Sciences* 43, e53. <https://doi.org/10.1017/S0140525X19002164>