enough to arouse any response beyond a depiction, an imitation of something that has agency. But as artificial agents gain in sophistication and intelligence, it is likely that humans will treat them as having real agency.

Competing interest. None.

References

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, 473–480. https://doi.org/10.1016/j.chb.2015.02.029

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*(4), 313–331. https://doi.org/10.1016/j.evolhumbehav.2004.08.015

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. http://www.jstor.org/stable/3328150

DeSilva, J. M., Traniello, J. F., Claxton, A. G., & Fannin, L. D. (2021). When & why did human brains decrease in size? A new change-point analysis & insights from brain evolution in ants. *Frontiers in Ecology and Evolution*, *9*, 742639. https://doi.org/10.3389/fevo.2021.742639

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franck, N. (2003). Awareness of action in schizophrenia. *Neuroreport*, *14*(7), 1081–1085. https://doi.org/10.1097/01.wnr.0000073684.00308.c0

Kaplan, M. (2022). After Google chatbot becomes "sentient," MIT professor says Alexa could too. *New York Post*, Retrieved from https://nypost.com/2022/06/13/mit-prof-says-alexa-could-become-sentient-like-google-chatbot/

Malafouris, L. (2008). Beads for a plastic mind: The "blind man's stick" (BMS) hypothesis & the active nature of material culture. *Cambridge Archaeological Journal*, *18*(3), 401–414. https://doi.org/10.1017/S0959774308000449

Malafouris, L. (2020). Thinking as "thinging": Psychology with things. *Current Directions in Psychological Science*, *29*(1), 3–8. https://doi.org/10.1177/0963721419873349

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*(2), 79–86. https://doi.org/10.1016/j.tics.2003.12.008

Ruff, C. B. (2005). Mechanical determinants of bone form: Insights from skeletal remains. *Journal of Musculoskeletal & Neuronal Interactions*, *5*(3), 202–212..

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science (New York, N.Y.)*, *333* (6043), 776–778. https://doi.org/10.1126/science.1207745

Stibel, J. M. (2021). Decreases in brain size & encephalization in anatomically modern humans. *Brain, Behavior and Evolution*, *96*(2), 64–77. https://doi.org/10.1159/000519504

van den Heiligenberg, F. M., Orlov, T., Macdonald, S. N., Duff, E. P., Henderson Slater, D., Beckmann, C. F., … Makin, T. R. (2018). Artificial limb representation in amputees. *Brain*, *141*(5), 1422–1433. https://doi.org/10.1093/brain/awy054

van den Heiligenberg, F. M., Yeung, N., Brugger, P., Culham, J. C., & Makin, T. R. (2017). Adaptable categorization of hands and tools in prosthesis users. *Psychological Science*, *28*(3), 395–398. https://doi.org/10.1177/0956797616685869

Social robots and the intentional stance

Walter Veit[a] and Heather Browning[b]

[a]School of History and Philosophy of Science, The University of Sydney, Sydney, NSW 2006, Australia and [b]London School of Economics and Political Science, Centre for Philosophy of Natural and Social Science, Houghton Street, London WC2A 2AE, UK
wrwveit@gmail.com; https://walterveit.com/
DrHeatherBrowning@gmail.com; https://www.heatherbrowning.net/

doi:10.1017/S0140525X22001595, e47

Abstract

Why is it that people simultaneously treat social robots as mere designed artefacts, 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.

Clark and Fischer (C&F) offer an excellent analysis of what they call the *social artefact puzzle*, that is, 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 artefacts, 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 artefacts, 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, 2022; Veit et al., 2019). 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 artefacts 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 nonhuman-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, Marchesi, & Wykowska, 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 artefact 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 C&F'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 & Wykowska, 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.

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

Competing interest. None.

References

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.

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., Dołega, 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

Binding paradox in artificial social realities

Kai Vogeley[a,b]

[a]Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, 59037 Cologne, Germany and [b]Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), 52428 Jülich, Germany
kai.vogeley@uk-koeln.de
k.vogeley@fz-juelich.de
https://psychiatrie-psychotherapie.uk-koeln.de/forschung/ag-soziale-kognition/
https://www.fz-juelich.de/de/inm/inm-3/forschung/soziale-kognition

doi:10.1017/S0140525X22001467, e48

Abstract

The relation between communication partners is crucial for the success of their interaction. This is also true for artificial social agents. However, the more we engage in artificial relationships, the more we are forced to regulate and control them. I refer to this as binding paradox. This deserves attention during technological developments and requires professional supervision during ongoing interactions.

Complementary to the technological development of artificial social agents, the question of how we can understand and conceptualize them in order to successfully communicate must be answered at the same time. This is the well-chosen focus of the target article by Clark and Fischer (C&F). They provide many examples for the different realizations of such agents (target article, sect. 3.2). That the relationship between two communication partners is crucial has been emphasized since the beginnings of modern social psychology (Watzlawick, Beavin, & Jackson, 1967).

In communication, we exchange information by conveying meaningful messages. According to symbolic interactionism, we interact on the basis of interpretable meanings that develop during the interaction between persons and can change over time (Blumer, 1969; Carey, 2009; Mead, 1963). However, content can only be transmitted if the communication partner is experienced as reliable and trustworthy. The "connectedness" or "attunement" between both partners is also referred to as rapport based on mutual attentiveness, reciprocal exchange of positivity cues, and coordination of nonverbal behaviors (Bernieri et al., 1996; Tickle-Degnen & Rosenthal, 1990). The relationship is the primary aspect of communication, while the content is secondary. For this reason, we tend to constantly interpret even unintended signals as meaningful: "we can not not communicate" (Watzlawick et al., 1967). These processes of communication do not always and necessarily occur unconsciously and