Note: This article has been published in a revised form in *Behavioral and Brain Sciences* [https://doi.org/10.1017/S0140525X22001455].

Target article: Clark & Fischer: Social robots as depictions of social agents
Title (Commentary): **Taking a Strong Interactional Stance**
Full Names: **Frank Förster** (1), **Frank Broz** (2), **Mark Neerincx** (3)
Institutions:
(1) Adaptive Systems Research Group, University of Hertfordshire
(2,3) Interactive Intelligence Research Group, Delft University of Technology
Mailing Addresses:
(1) Adaptive Systems Research Group, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK
(2,3) EEMCS, Building 28, Van Mourik Broekmanweg 6, 2628 XE Delft, NL
Email Addresses:
(1) f.foerster@herts.ac.uk, (2) F.Broz@tudelft.nl, (3) mark.neerincx@tno.nl
Home Page URLs:
(1) https://frank-foerster.gitlab.io
(2) https://www.tudelft.nl/en/eemcs/the-faculty/departments/intelligent-systems/interactive-intelligence/people/current-group-members/frank-broz
(3) https://www.tudelft.nl/ewi/over-de-faculteit/afdelingen/intelligent-systems/interactive-intelligence/people/current-group-members/mark-a-neerincx

Abstract
We outline two points of criticism. Firstly, we argue that robots do constitute a separate category of beings in people's minds rather than being mere depictions of non-robotic characters. Secondly, we find that (semi-)automatic processes underpinning communicative interaction play a greater role in shaping robot-directed speech than Clark's and Fischer's theory of social robots as depictions indicates.

Main text:
We formulate two points of criticism regarding Clark's & Fischer's contribution and suggest that common research practices in human-robot interaction contribute to reinforcing confusion about robot capabilities by obfuscating the nature of the interaction with an agent or prop.
Firstly, we argue that robots do exist as a separate class of entity in people's minds even before they encounter an actual robot in real life. This mental model that varies amongst people is likely due to their exposure to fictional depictions of robots in popular media. People know and expect that a robot dog or a humanoid robot is a different kind of entity than a dog or a person. They are unclear on the actual capabilities of these agents, but they can and will discover this through interaction, which makes robots distinct from noninteractive depictions such as static art or characters in noninteractive performances. Research methodology in human-robot interaction, e.g. a widespread use of wizard of oz experimental designs, and a lack of transparency about the level of a robot's autonomy reinforces this ambiguity about capabilities. Fischer and Clark present a virtual agent or a ventriloquist's dummy as similar examples of agents. But we argue that these agents engage in very different types of interactions, where in one case the agent being

interacted with is an autonomous computer program and in the other the interaction is with another person through the use of a prop with the human controlling this prop being visible and known to their interaction partner.
Secondly, Clark and Fischer underplay the influence of (semi-) automatic processes on the concrete trajectory and form of an interaction due to this conflation of interactive and noninteractive formation of understanding of agents or characters. While a person's speech style initially may be influenced by depictions as construed by the authors, the affordances and real-time contingencies of the unfolding interaction will substantially impact upon that person's style of talk. Some of these real-time adaptations are automatic (such as gaze in face-to-face conversation, Broz et al. 2012) and may "pull" the unfolding interaction in a direction different to the one set up by the person's pre-existing views of the robot's role or nature.

In support of this view are the following transcripts originating from the negation acquisition studies conducted by Förster et al. (2019). These studies consisted of multiple sessions per participant, and the transcripts pertains both to participant P12 (P) teaching object labels to Deechee (D), a childlike humanoid robot that was presented to participants as a young language learner.

Session 2, 0 min 47 seconds
((P picks up heart object))
P-1 *this one here is a heart*
P-2 *you don`t like the shape*
((P turns object around))
P-3 *do you wanna see upside down*
P-4 *heart*
((D turns head and frowns))
P-5 *no don`t like that [one]*

Session 5, 1 min 20 seconds
((P picks up square, D reaches out for it))
D-1 *square!*
((D gets to hold object and drops it))
P-6 *yeah square!*
((P picks up triangle))
D-2 *done!*
P-7 *no! (0.5s) don't say done!*
D-3 *triangle*
P-8 *yeah, triangle! (..) [well done!]*
[((P puts down triangle))
((P picks up heart))
D-4 *done*
P-9 *no, I say well done (.) you don't say done*
P-10 *what's this one?*
D-5 *heart!*
P-12 *yes*

Participant P12, instructed to talk to Deechee as if it was a 2-year old child, initially spoke in a style roughly compatible with child-directed speech. This included intent-related questions (P-3) and intent interpretations (P-2, cf. Förster et al. 2018). During the second session, however, P12 decided to speak in a much simpler, “robotic” register, that he maintained during the two follow-up sessions and into his fifth session. In this register he used mostly one-word utterances that consisted either of object labels or short feedback words, e.g. P-6 and P-8. This change, as we learned later, was meant to optimize the learning outcome of the - by him - hypothesised learning algorithm such that his mental model of the robot was arguably one of a mere machine. However, once Deechee started to use negation words such as `*no*' or `*done'* (D-2 and D-4), P12 did not manage to maintain his linguistic restraint and abandoned his minimalistic speech style for short time periods (e.g. P-7 and P-9).
Given P12’s strong adherence to his chosen minimalistic speech register prior to these lapses, these utterances appear to have a somewhat involuntary character. We argue that these lapses were caused by automatic processes temporarily gaining the upper hand over the conscious, self-imposed restrictions. The “pull” of the interaction caused the participant to treat Deechee, at least temporarily, as a being with wants or emotions. This change is due to Deechee’s behaviour-in-interaction rather than a unilateral perspective switch in terms of class of depiction (cf. Förster & Althoefer, 2021). In terms of being seen as a depiction of another character it is unclear what that could possibly be in this setting. Deechee does not serve any distinct social role such as receptionist nor does it correspond to a known character such as Kermit the frog.

For social robots to be useful in their intended roles, they must become (and be understood as) social agents in and of themselves rather than puppets that experimenters act through to investigate people’s incorrect mental models. This will necessarily involve people coming to understand their capabilities and limitations through multiple and prolonged interactions. More generally, the application of data-driven machine learning technology in successive human-robot collaborative activities will involve co-adaptation and co-learning. Such new emergent behaviours may comprise unconscious tangible interactions (Zoelen et al., 2021a) and new collaboration patterns (Zoelen et al., 2021b). This way, the human develops cognitive, affective and tangible experiences and understandings of the robots, grounded in the pursuing situated collaborations. In addition to the “pre-baked” designs (Ligthart et al., 2019), anthropomorphic projections (Carpenter, 2013) and human-like collaboration functions (Neerincx et al., 2019), the evolving unique robot features with corresponding behaviours will affect the continuous (re-)construction of new types of robot characters.

Acknowledgements

Conflict ot Interest Statement
The authors declare no conflict of interest.
Funding Statement
The cited transcripts originate from work that was supported by the EU Integrated Project “Integration and Transfer of Action and Language in Robots” through the European Commission under Contract FP-7-214668

Reference List

Broz, F., Lehmann, H., Nehaniv, C. L., and Dautenhahn, K. Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation,2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, 2012, 858-864, https://doi.org/10.1109/ROMAN.2012.6343859

Carpenter, J. (2013). The Quiet Professional: An investigation of US military Explosive Ordnance Disposal personnel interactions with everyday field robots [Doctoral dissertation, University of Washington]. ResearchWorks Archive. http://hdl.handle.net/1773/24197

Förster, F., Althoefer, K. (2021). Attribution of autonomy and its role in robotic language acquisition. *AI & Society*. 37(2), 605-617, https://doi.org/10.1007/s00146-020-01114-8

Förster, F., Saunders, J., Lehmann, H., and Nehaniv, C. L. (2019). Robots Learning to Say “No”: Prohibition and Rejective Mechanisms in Acquisition of Linguistic Negation. *ACM Transactions on Human-Robot Interaction*, *8*(4), 1-26. https://doi.org/10.1145/3359618

Förster, F., Saunders, J., and Nehaniv, C. L. (2018). Robots That Say “No”: Affective Symbol Grounding and the Case of Intent Interpretations, *IEEE Transactions on Cognitive and Developmental Systems*. *10*(3), 530-544. https://doi.org/10.1109/TCDS.2017.2752366

Ligthart, M., Fernhout, T., Neerincx, M. A., van Bindsbergen, K. L., Grootenhuis, M. A., & Hindriks, K. V. (2019). A Child and a Robot Getting Acquainted - Interaction Design for Eliciting Self-Disclosure. In Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems, 61-70.

Neerincx, M. A., Van Vught, W., Blanson Henkemans, O., Oleari, E., Broekens, J., Peters,

R., Kaptein, F., Demiris, Y., Kiefer, B., Fumagalli, D., Bierman, B. (2019). Socio-Cognitive Engineering of a Robotic Partner for Child's Diabetes Self-Management. *Frontiers in Robotics and AI*, *6*. https://doi.org/10.3389/frobt.2019.00118

Van Zoelen, E. M., Van Den Bosch, K., & Neerincx, M. (2021). Becoming Team Members: Identifying Interaction Patterns of Mutual Adaptation for Human-Robot Co-Learning. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.692811

Van Zoelen, E. M., Van Den Bosch, K., Rauterberg, M., Barakova, E., & Neerincx, M. (2021). Identifying Interaction Patterns of Tangible Co-Adaptations in Human-Robot Team Behaviors. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.645545