This is an accepted manuscript version of the paper published in:

Wykowska, A. (2021). Robots as mirrors of the human mind. Current Directions on Psychological Science, 30, 34-40. doi: https://doi.org/10.1177/0963721420978609

Corresponding Author:

Agnieszka Wykowska

Istituto Italiano di Tecnologia

Social Cognition in Human-Robot Interaction Unit

via Enrico Melen 83

16152 Genova, Italy

Phone: +39 010 8172242

Email: agnieszka.wykowska@iit.it

Word count: 3014 (including abstract and keywords, and acknowledgment, but excluding list of references, list of recommended readings and figure captions)

-- Title page --

Title: Robots as mirrors of the human mind

Author:

Agnieszka Wykowska,

Istituto Italiano di Tecnologia,

Social Cognition for Human-Robot Interaction Unit

Genova, Italy

-- Abstract page --

Abstract: Robots are currently the center of attention in various fields of research, due to their potential use as assistants for daily living. In this paper, I highlight a different role that robots can also play, namely being a tool for understanding human cognition. I provide examples where robots have been used in experimental psychology to study socio-cognitive mechanisms such as joint attention or sense of agency. I also discuss the issue of whether and when robots (especially those that resemble humans) are perceived through a human-centered lens, with anthropomorphic attributions. In the final section, I describe approaches in which the robots' embodiment is used for implementation of computational models of human cognition. In sum, the collection of studies presented here shows that robots can be an extremely useful tool for scientific inquiry in the area of experimental psychology and cognitive science.

Keywords: Human-robot interaction, Robots in experimental psychology, Social cognition.

-- Main text --

1. Introduction

Humanoid robots are often associated with science fiction movies and futuristic imageries, think of C-3PO or Commander Data, for example. However, also in science, humanoid robots, equipped with powerful modern AI, are receiving substantial attention. It is predicted that robots will soon enter our daily life as assistants in workplace, shops, airports or hospitals; as educators or as social companions in healthcare or elderly care (see Royakkers & van Est, 2015 for a review on the roles of robotics in present and future society and related ethical considerations).

In this paper, I highlight a different role that robots can also play, especially in the context of basic scientific research, namely serving as tools for examining our own human nature. Robots can be informative with respect to understanding the human mind, as they can be tools to study social cognition by means of their embodied physical (and perhaps even social) presence, which provides higher ecological validity than screen-based stimuli and better experimental control than human-human interaction. By using robots as sophisticated stimuli, we can:

(1) learn whether cognitive mechanisms observed in classical experiments with 2D stimuli on the screen that have been identified as mechanisms enabling interaction with other humans also generalize to interactive and more naturalistic scenarios with artificial agents,

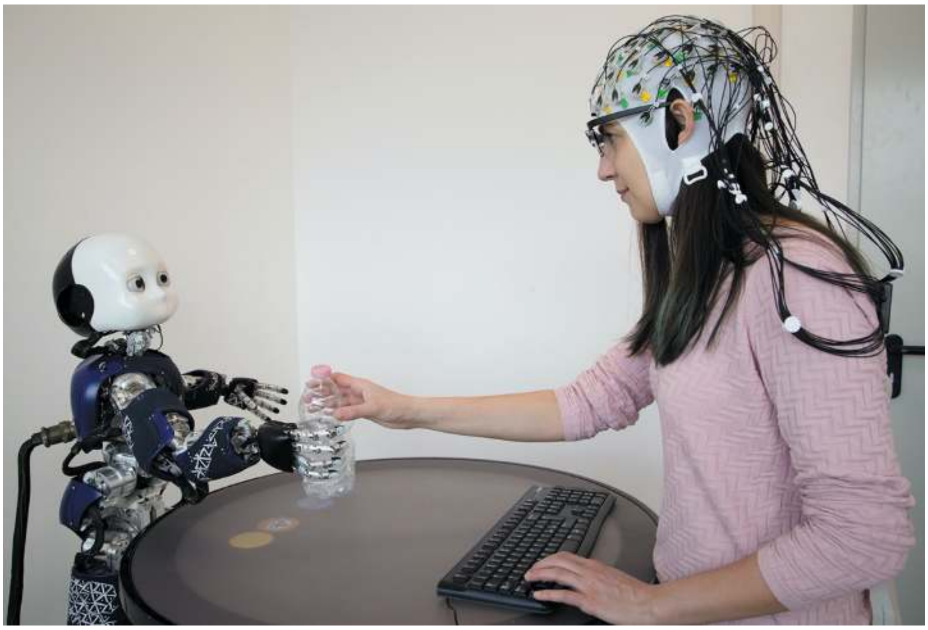
(2) study phenomena that would otherwise be difficult to examine due to practical or technical limitations,

(3) examine what constitutes humanness.

Apart from the benefits that robots can provide in the role of sophisticated stimuli, they can also serve the role of embodied computational models of cognition. As such, they could potentially be a unique tool which allows generating new hypotheses, predictions and mechanistic explanations regarding human cognition.

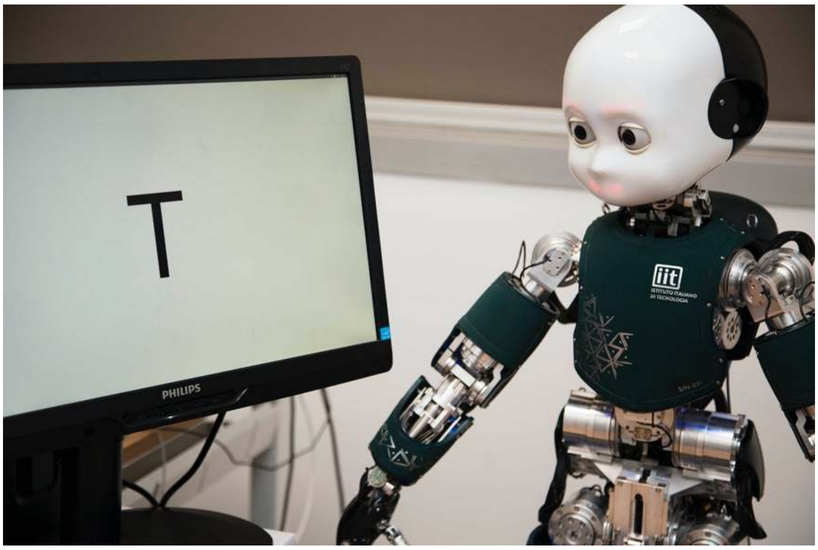
2. Robots as sophisticated stimuli: from 2D faces on the screen to embodied physical presence

The question of how mechanisms of human social cognition work has been addressed in psychological research with the use of strictly controlled types of stimuli, such as schematic faces presented on a screen in a highly controlled lab environment. This more “classical” approach has been criticized for lack of ecological validity, which, in the case of social cognition, is particularly crucial. Recent approaches, motivated by the second-person neuroscience framework of Schilbach et al. (2013), underline that understanding the mechanisms of social cognition requires interactive protocols where participants are engaged in reciprocal interaction, rather than being only in the spectatorial mode passively observing stimuli on the screen. However, adopting more ecologically valid experimental protocols, with natural interactions between humans, encounters the challenge of reduced experimental control. Therefore, approaches using virtual reality and virtual avatars have been developed (e.g., Pan & Hamilton, 2018). However, virtual avatars lack physical embodiment, which might also affect social cognition and engagement (for review, see Li, 2015). In fact, there is evidence that physical presence and embodiment is crucial for various cognitive mechanisms and learning (Roseberry Lytle, Garcia-Sierra, Kuhl, 2018), although the exact reasons are yet to be examined. Importantly, physical embodiment and presence entails sharing space in the environment and potential for manipulating the shared environment. In this context, robots might prove beneficial, as they offer physical embodied presence on the one hand, and possibility of experimental control on the other, cf. Figure 1.



-- Figure 1 --

An excellent example of examining social cognition with the use of a humanoid robot's physical presence is studying joint attention, one of the most fundamental mechanisms, on which higher-level socio-cognitive processes are based (Baron-Cohen, 1995). Joint attention occurs when one interaction partner directs his/her attention to a location or event in the environment, and the other attends there in response. In laboratory settings, joint attention has been operationalized, for example, in form of a gaze-cueing protocol (Friesen & Kingstone, 1998) where directional shifts of gaze elicit attentional orienting in an observer. In a typical gaze-cueing protocol a face or face-like stimulus is presented on the screen either with eyes directed straight-ahead, or with empty placeholders in the location of the eyes, in case of schematic face drawings. After a while, the gaze direction changes towards a location on the screen. Subsequently, a target stimulus is presented either at the gazed-at location or elsewhere. Typically, performance related to the target (reaction times, or error rates, in target detection or discrimination) is better for gazed-at locations, as compared to the other locations, an effect that is postulated to manifest gaze-induced attentional orienting.



-- Figure 2 --

Kompatsiari and colleagues (Kompatsiari et al., 2018a) implemented a gaze cueing study in a protocol involving a physically embodied humanoid robot iCub (Metta et al., 2010) and stimuli presented on two laterally positioned screens. The authors observed standard gaze cueing effects both at the behavioral and neural level, the latter in the form of P1-N1 complex of event-related potentials (ERPs) of the EEG signal. These findings demonstrate that results obtained with 2D stimuli on the screen generalize to physically present embodied robots, and that such robots are capable of inducing joint attention in general.

Similarly, Pérez-Osorio et al. (2018) implemented a paradigm from earlier studies (Pérez-Osorio, 2017) in a human-robot interaction protocol. In this paradigm, gaze cueing has been embedded in an action context. The results showed that gaze cue-elicited attentional orienting is dependent on action expectations of participants regarding the successive action steps of the robot, thereby replicating previous findings (Pérez-Osorio, 2017) and suggesting that similar attentional mechanisms are at stake when interacting with a robot as they have been observed with 2D stimuli presented on the screen. This study, similarly to a growing body of studies using physical interaction to study coordination and synchronization (e.g., Mörtl, Lorenz & Hirche, 2014), shows the advantage of using embodied physically present robots for measuring perceptual, cognitive and motor processes in naturalistic interactive scenarios.

Another interesting mechanism to study in the social context is sense of agency. It has been shown that humans' sense of agency, in the sense of perceived control over the outcomes of one's own actions (Gallagher, 2000), is affected by the presence of others. With the use of 2D avatars on the screen, Beyer and colleagues found that sense of agency can be reduced in a social context – a phenomenon that has been argued to underlie the "bystander effect" or diffusion of responsibility (Beyer et al., 2017). Ciardo et al. (2020) adapted Beyer et al.'s paradigm to a human-robot interaction study and showed that humans – when engaged in a game with a robot, in which action (or inaction) can be very costly – experience a reduced sense of agency, as compared to performing the task alone. This effect was observed also in interaction with another human, but not with a purely mechanical device (an air pump). Thus, once again, phenomena observed earlier with the use of stimuli on the screen have been shown to generalize to more natural interaction protocol.

3. Robots as a technological solution for studying socio-cognitive phenomena that are difficult to examine with other methods

Kompatsiari and colleagues examined the impact of mutual gaze on the mechanisms of joint attention (Kompatsiari et al., 2018b) with the use of the iCub robot. Examining the impact of mutual gaze on joint attention would be technically difficult (and unnatural) with stimuli presented on the screen. On the other hand, examining mutual gaze in human-human interaction is challenging due to maintaining high degree of experimental control, i.e., the difficulty of implementing a repeated identical behavior (for example, mutual vs. aversive gaze) over many trials, with the same timing parameters and exactly identical (controlled) movement of the eyes and neck. In this context, a robot stimulus offers an excellent opportunity of ecological validity (natural movements of mechanical eyes, combined with an algorithm for detecting the observer's eyes and engaging in actual, real-time mutual gaze) and experimental control (identical movement over many trials).

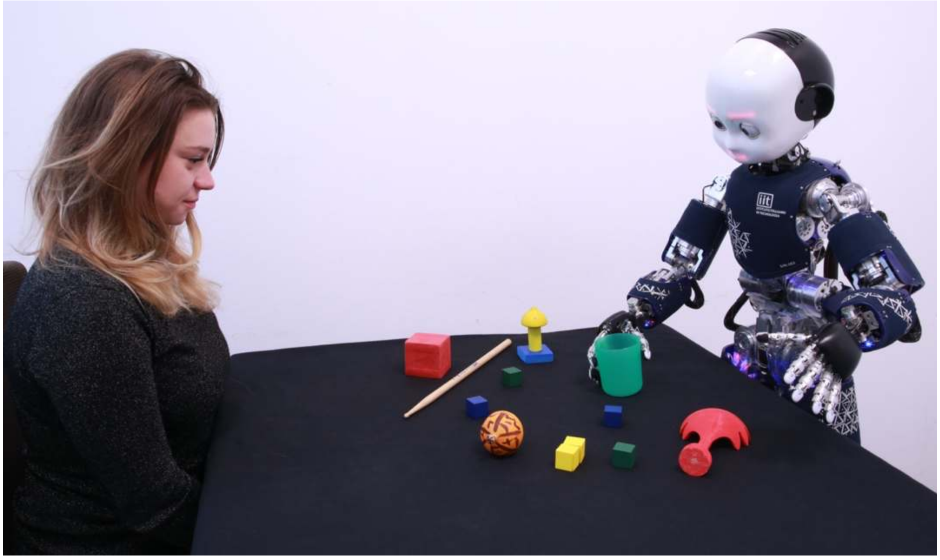
In Kompatsiari et al.'s study (2018b) iCub engaged participants in direct eye contact, or looked elsewhere, before the gaze cueing procedure (Figure 2). The results showed that the mechanism of attentional orienting relies on a complex interplay between various attentional and social mechanisms, as mutual gaze modulated the magnitude of gaze cueing. Furthermore, the results also showed that the robot's direct "gaze" exhibited by two eye-like cameras was interpreted in a social manner, which participants found engaging (Kompatsiari et al., 2018b). As argued above, such findings would be difficult to obtain with either screen-based stimuli or natural human-human interaction.

Willemse & Wykowska (2019) developed a protocol in which the iCub robot was the follower in joint attention, and the likelihood of it following participants' gaze was manipulated (participants were introduced to the robot "Jimmy" or "Dylan", where Jimmy followed their gaze in 80% of trials, while Dylan only in 20% of trials). The results showed that people re-engage with the robot face faster when it follows their directional gaze, and that robot "identity", which followed participants' gaze with higher likelihood, was liked more. This demonstrates that the human socio-cognitive system is sensitive to reciprocal behavior of others, and artificial agents might actually be more engaging and treated more socially if they exhibit a certain degree of reciprocity. Similar to the previous example of Kompatsiari and colleagues' studies, using a robot in this study allowed for manipulating contingent behavior very precisely and in a well-controlled manner over many trials – a manipulation that would be difficult in a naturalistic interactive setup with another human agent.

4. Robots to inform us about what constitutes humanness

One interesting question to ask is whether robots can tell us what "being human" means to us. What are the truly specific features of humanness? Can robots "pretend" humanness just by physically resembling humans, like the famous Ishiguro's androids (Ishiguro, 2006)? Attribution of humanness to artificial agents has been extensively studied in the scientific field of human-robot interaction. Several authors have identified factors contributing to anthropomorphism in general (e.g., Epley, Waytz, & Cacioppo, 2007) and in human-robot interaction specifically (Złotowski et al., 2015) pointing out internal cognitive factors of the observer, such as accessibility to knowledge about an agent, motivation to explain the behavior of others, and desire for social contact (e.g., Epley, Waytz, & Cacioppo, 2007) as well as factors related to the observed agent, such as human-like appearance or behavior (Fink, 2012, Złotowski et al., 2015).

In the context of anthropomorphism, it is essential to highlight that the core aspect of humanness is endowment with unobservable internal mental states. Do we potentially attribute mental states to artificial agents, such as computer programs or robots? Several authors empirically addressed the issue of attribution of mind to various entities (e.g., Gray, Gray & Wegner, 2007) or more specifically, mental states to artificial agents (e.g., Chaminade et al., 2012; Thellman, Silvervarg, & Ziemke, 2017). Marchesi and colleagues (2019) addressed this question with their Intentional Stance Questionnaire, in which participants were asked to choose a description that fits best to the observed storyline involving the iCub robot. Some descriptions were using mentalistic vocabulary (which would assume some degree of attribution of mental states to robot) and some involved only mechanistic words. The results showed that participants were more likely to use mechanistic terms in describing robot behaviors, but the mentalistic explanations were not uncommon, and some people were more likely to use them than others. Interestingly, in a follow-up study, Bossi et al. (under review) showed that the individual “bias”, that is, the likelihood of choosing more mentalistic or more mechanistic explanations of robot behaviors, can be discriminated already from the resting state EEG signal, meaning signal measured before participants even perform the task or are exposed to the robot whose behavior they later need to describe.



-- Figure 3 --

Thellman et al. (2017) showed similar results to Marchesi and colleagues, indicating that people use similar descriptions for robot behavior as they would use for humans, specifically regarding ascription of intentionality.

De Graaf & Malle (2019) presented participants with text descriptions of robot actions, without presenting visual representation of any robot embodiment. Participants were asked to provide free-response verbal explanation of a given behavior. The explanations were then analyzed in terms of reasons and mental states (belief and desire reasons), and the results showed that humans used similar explanations of behavior for robots as they did for humans.

These examples of studies demonstrate that humans sometimes do use mentalistic reasoning to explain robot behavior. However, it might be the case that attributing to artificial agents mental states of non-phenomenal, “cognitive”, nature (e.g., beliefs) is more likely than phenomenal (e.g., pain) or affective (e.g., happiness) states (Huebner, 2010).

Overall, it seems that our cognitive system sometimes uses the tools it has developed for interacting with conspecifics for new situations, such as social interactions with artificial agents. This tells us something about ourselves – sometimes we use our old and comfortable ways of thinking about other humans (agents with which we have most expertise and exposure) for novel categories of entities. However, as our knowledge about the workings of the observed system becomes available, we might switch from the more accessible anthropomorphic reasoning, with which we have most experience, to the alternative way of explaining other entities or agents (Epley, Waytz, & Cacioppo, 2007). Interestingly for the purposes of future research, the determinants of inclusion of robots into the human-like category of intentional agents still remain to be answered, as there is also evidence that humans might have negative attitudes or anxiety towards robots (for a detailed meta-analysis of literature on positive and negative attitudes towards robots, see Naneva et al., 2020). In this context, it seems plausible that people might treat robots as out-group members, as some of the items on the most-commonly used questionnaire developed to measure negative attitudes (Nomura et al., 2006) address in-group/out-group membership (e.g., items "I surmise that something negative for humans happen when robots become more similar to humans" or "I feel anxiety if robots really have their own emotions"). This idea would also be in line with the finding of De Graaf et al. (2016) showing stronger associations of negative words with robots and positive words with humans in an implicit association test. A more specific question that still needs to be addressed is what are the exact factors that play a role in attribution of mental states to other agents, although some factors have already been identified in literature (Epley, Waytz, & Cacioppo, 2007; Złotowski et al., 2015).

5. Robots as embodied computation models of human cognition

Robots can also constitute means for understanding human cognition and human cognitive development (Asada et al., 2009) by serving as embodied models of the cognitive mechanisms, or, in other words, "understanding through building" (Verschure & Prescott, 2018). In this vein, Prescott and colleagues (2019) implemented on the iCub robot a multimodal memory system grounded in theories and models from cognitive psychology and neuroscience and showed that the model allows better social functioning of the robot.

Morse and colleagues (Morse at al., 2015) proposed a neural network architecture for acquisition of word-object mapping within the embodied cognition framework. Their architecture was implemented on the iCub robot. The architecture consisted of three "fields" (pools of nodes): the "visual field", the "body posture field" and the "word field". The important feature of the architecture was that binding of visual representation of objects (visual field) with words (word field) occurred through body-centric representation of space.

In a series of experiments, Morse and colleagues (2015) examined the role of spatial consistency between visual object presentation and auditory presentation of the object label in word-object mapping, as spatial consistency has been found to impact infants' learning of the object-name associations. Importantly, however, in Morse and colleagues' architecture, the spatial representation was body-centric, through representation of the robot's joint angles of eyes, head and torso. The authors showed that the robot was able to learn word-object mapping through links via the body posture representation, and that the mapping is generalizable to new spatial configurations. The results showed also interference effects when the spatial representation was disrupted during learning, either by lack of consistent mapping between body posture and object identity or by changes in body posture during the naming event. Importantly, the model generated predictions regarding human performance in an analogous task, and those predictions have been confirmed in a subsequent series of experiments with infants. In sum, the model of Morse and colleagues confirmed the importance of spatial information during acquisition of object naming and suggested that the spatial representation might actually be body-centric. Importantly for the purposes of this paper, this approach showed that computational models implemented in a robot embodiment can generate new theoretical predictions and inspire new research to test the predictions.

6. Concluding remarks

In this paper, I aimed to show that robots, employed as sophisticated "stimuli", can inform us about our own cognitive mechanisms, or, in the role of embodied computational models, can generate new theoretical predictions regarding the workings of the human brain. Therefore, robots' beneficial role for humans extends beyond being assistants and companions in various domains of human life. They can be a very useful tool for the basic research in experimental psychology and cognitive science.

7. Acknowledgement

This work has received support from the European Research Council under the European Union's Horizon 2020 research and innovation programme, G.A. ERC-2016-StG-715058.

References

1. Asada M. et al., (2009). Cognitive Developmental Robotics: A Survey. IEEE Transactions on Autonomous Mental Development, 1, 12-34
2. Baron Cohen, S. (1995) Mindblindness: Essay on Autism and the Theory of Mind. *MIT Press*, Cambridge, MA.
3. Bossi, F., Willemse, C., Cavazza, J., Marchesi, S., Murino, V., & Wykowska, A. (in review). EEG activity pattern during resting state predicts biases in attitudes towards robots, *Science Robotics*.
4. Beyer, F., Sidarus, N., Bonicalzi, S., Haggard, P. (2017). Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring. *Social Cognitive and Affective Neuroscience, 12*, 138-145.
5. Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience, 6:103*.
6. Ciardo, F., Beyer, F., De Tommaso, D., Wykowska, A. (2020). Attribution of intentional agency towards robots reduces one's own sense of agency. *Cognition, 194: 104109*.
7. De Graaf, M. M. A., Malle, B. F. (2019). People's explanations of robot behavior subtly reveal mental state inferences. In: ACM/IEEE International Conference on Human-Robot Interaction, *IEEE Xplore, doi:* 10.1109/HRI.2019.8673308.
8. De Graaf, M. M. A., Allouch, S. B., Lutfi, S. (2016). What are people's associations of domestic robots? Comparing implicit and explicit measures. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, 1077–1083.
9. Epley, N., Waytz, A., Cacioppo, J.T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review, 114*, 864–886.
10. Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-

robot interaction. *Lecture Notes in Artificial Intelligence, 7621*, 199–208.

11. Friesen, C.K., Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review, 5*, 490-495.
12. Gallagher, S. (2000). Philosophical conception of the self: implications for cognitive science. *Trends in Cognitive Science*, *4*, 14-21.
13. Gray, H. M., Gray, K., Wegner, D. M. (2007). Dimensions of mind perception. Science 315 (5812): 619.
14. Huebner, B. (2010). Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences, 9*, 133–155.
15. Ishiguro, H. (2006). Android science: Conscious and subconscious recognition. *Connection Science, 18*, 319-332.
16. Kiesler, S., Powers, A., Fussell, S.R., Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition, 26*,169–181.
17. Kompatsiari, K., Perez-Osorio, J., De Tommaso, D., Metta, G., Wykowska, A. (2018a). Neuroscientifically-Grounded Research for Improved Human-Robot Interaction. In: IEEE International Conference on Intelligent Robots and Systems, *IEEE Xplore, doi:* 10.1109/IROS.2018.8594441.
18. Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., Wykowska, A. (2018b). On the role of eye contact in gaze cueing. *Scientific Reports, 8*: *17842*.
19. Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies, 77*, 23-37.
20. 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*.
21. Metta, G., et al. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks, 23*,1125-1134.
22. Morse, A.F., Benitez, V.L., Belpaeme, T., Cangelosi, A., Smith, L.B. (2015). Posture affects how robots and infants map words to objects. *PLoS One, 10(3): e0116012*.
23. Mörtl, A., Lorenz, T., Hirche, S. (2014). Rhythm Patterns Interaction - Synchronization Behavior for Human-Robot Joint Action. *PLoS ONE 9(4): e95195*.
24. Naneva, S., Sarda Gou, M., Webb, T.L. & Prescott, T. (2020). A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots. *International Journal of Social Robotics*. Online first: https://doi.org/10.1007/s12369-020-00659-4.

25. Nomura, T., Kanda, T., Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI & Society, 20,* 138–150.
26. Pan, X., Hamilton, A.F.C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*, 395–417.
27. Pérez-Osorio, J., De Tommaso. D., Baykara, E., Wykowska, A. (2018). Joint Action with iCub: A Successful Adaptation of a Paradigm of Cognitive Neuroscience in HRI. In: RO-MAN IEEE International Symposium on Robot and Human Interactive Communication, *IEEE Xplore,* doi: 10.1109/ROMAN.2018.8525536.
28. Pérez-Osorio, J., Müller, H., Wykowska, A. (2017). Expectations regarding action sequences modulate electrophysiological correlates of the gaze-cueing effect. *Psychophysiology, 942-954.*
29. Prescott, T. J., Camilleri, D., Martinez-Hernandez, U., Damianou, A., Lawrence, N. D. (2019). Memory and mental time travel in humans and social robots. *Philosophical Transactions of the Royal Society. B: Biological Sciences, 374*, 20180025.
30. Roseberry Lytle, S., Garcia-Sierra, A., Kuhl, P. K. (2018). Infant language learning in the presence of peers. *Proceedings of the National Academy of Sciences, 115,* 9859-9866.
31. Royakkers, L., van Est, R. (2015). A Literature Review on New Robotics: Automation from Love to War. *International Journal of Social Robotics 7*, 549–570.
32. Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*, 393-414.
33. Thellman, S., Silvervarg, A., Ziemke, T. (2017). Folk-psychological interpretation of human vs. humanoid robot behavior: Exploring the intentional stance toward robots. *Frontiers in Psychology, 8*: *1962*.
34. Verschure, P.F.M.J., Prescott, T.J. (2018). A living machines approach to the sciences of mind and brain. In: Prescott, T.J., Lepora, N., Verschure, P.F.M.J. (eds.) The Handbook of living machines: research in biomimetic and biohybrid systems. *Oxford University Press*, Oxford, UK, pp. 15-25.
35. Willemse, C., Wykowska, A. (2019). In natural interaction with embodied robots, we prefer it when they follow our gaze: A gaze-contingent mobile eyetracking study. *Philosophical Transactions of the Royal Society. B: Biological Sciences, 374: 20180036.*
36. Złotowski, J., Proudfoot, D., Yogeeswaran, K., Bartneck, C. (2015) Anthropomorphism: opportunities and challenges in human–robot interaction. *International Journal of Social Robotics, 7*, 347–360.

-- Recommended readings page --

Cangelosi, A., & Schlesinger, M. (2015). Developmental Robotics: From Babies to Robots. The MIT Press, Cambridge, Massachusetts: A book providing an overview of the field of Developmental Robotics. The book provides an excellent introduction to how computational models, inspired by developmental psychology, learn and instantiate various cognitive mechanisms in embodied systems.

Korman, J., Voiklis, J., Malle, B. (2015). The social life of cognition. *Cognition, 135*, 30-35: An overview paper discussing human social cognition in the context of general cognitive mechanisms, and the debate between domain-specificity of social cognition, which concludes with an outlook into future directions and methods, where robots are presented as tools for embodying computational models of human cognition, and as tools to test social cognitive mechanisms such as ascription of intentionality or moral judgments.

Perez-Osorio, J. & Wykowska, A. (2020). Adopting the intentional stance toward natural and artificial agents, *Philosophical Psychology, 33*, 369-395: A review paper discussing the concept of intentional stance from philosophical and developmental perspective, addressing the issue of adopting intentional stance towards natural and artificial agents.

Wiese, E., Metta, G., Wykowska, A. (2017). Robots as Intentional Agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology, 8:1663*: A review paper arguing in favour of research approaches crossing the boundaries of disciplines and linking methods of social and cognitive neuroscience with robotics, in order to improve robot design on the one hand and understanding of human cognition on the other.

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society London: B. Biological Sciences, 371, 20150375*: A review paper describing in length how robots can be used as tools to understand human social cognition

-- Figure Captions --

Figure 1. An experimental setup where a natural interaction is introduced, and experimental control is maintained by means of experimental design and setup. During the interactive protocol cognitive processes can be measured not only with behavioral measures (reaction times and error rates) but also with EEG and eye tracking, as illustrated here.

Figure 2. iCub cueing participants' attention to the target letter by means of its eye and head direction

Figure 3. One example stimulus from the Intentional Stance Questionnaire (Marchesi et al., 2019), which participants observed, and for which they were asked to choose an explanation with either mentalistic or mechanistic vocabulary.