Intentional Mindset Toward Robots—Open Questions and Methodological Challenges

Elef Schellen and Agnieszka Wykowska*

Social Cognition in Human-Robot Interaction, Istituto Italiano di Tecnologia, Genoa, Italy

Natural and effective interaction with humanoid robots should involve social cognitive mechanisms of the human brain that normally facilitate social interaction between humans. Recent research has indicated that the presence and efficiency of these mechanisms in human-robot interaction (HRI) might be contingent on the adoption of a set of attitudes, mindsets, and beliefs concerning the robot's inner machinery. Current research is investigating the factors that influence these mindsets, and how they affect HRI. This review focuses on a specific mindset, namely the "intentional mindset" in which intentionality is attributed to another agent. More specifically, we focus on the concept of adopting the intentional stance toward robots, i.e., the tendency to predict and explain the robots' behavior with reference to mental states. We discuss the relationship between adoption of intentional stance and lower-level mechanisms of social cognition, and we provide a critical evaluation of research methods currently employed in this field, highlighting common pitfalls in the measurement of attitudes and mindsets.

Keywords: HRI (human robot interaction), intentional stance, humanoid robot, social cognition, social neuroscience

INTRODUCTION

Human-robot interaction (HRI) is an increasingly important topic, as robotic agents are becoming more developed, and are likely to take a conspicuous role as social agents in fields like health care and education, as well as daily living (Broadbent et al., 2009; Cabibihan et al., 2013). Humanoid robots, specifically, occupy a unique niche in our psychology (Wykowska et al., 2016; Wiese et al., 2017), as they are artificial human-like agents and thereby unlike any other natural social stimulus. Furthermore, they are distinctly different from other artificial agents such as virtual characters or avatars, in that they are embodied. Mapping out the way in which HRI differs from human-human interaction is essential for creating robots that successfully and efficiently interact with end-users. In this review, we will focus on a specific attitude or mindset that humans have toward other humans, namely the intentional stance (Dennett, 1987). We will consider the issue of adopting intentional stance toward artificial embodied agents (robots), we will discuss its relationship to other, lower-level mechanisms of social cognition, and we will critically evaluate methods to assess adoption of intentional stance. We posit that an intentional mindset, in which we frame a robot's actions in terms of its goals and desires, on the part of the user, might be crucial for well-functioning HRI.

INTENTIONAL STANCE IN HRI

Dennett (1971, 1987) has proposed a categorization of different kinds of attitudes that humans have when explaining and predicting behavior of an observed system. This categorization includes three *stances* of differing complexity: the physical stance, the design stance and the intentional stance. These stances are used to heuristically make sense of events around us. In the *physical stance*, events are explained using physical rules of cause and effect; we know that water will boil when sufficiently heated, and that a ball rolling toward the edge of a table will fall down after crossing it. Using the *design stance*, one explains occurrences based on the assumption that these are outcomes of a purposefully designed system. This stance makes for an expedient framework facilitating interaction with systems that have input-output rules, like operating an industrial machine, where certain button presses correspond with certain machine operations. Finally, the *intentional stance* rests on the assumption that beliefs and desires underlie behavioral outcomes. Social robots occupy an ambiguous position in this framework, as they do not possess beliefs and desires in the folk psychology interpretation of these concepts. Still, they may be programmed to act as if they do, and adopting the intentional stance may result in the consequent interaction being more efficient and pleasant. The intentional stance, as it applies to HRI, shares many similarities with the concept of anthropomorphism. When a user ascribes human motivations and desires to a robot, in interpreting and predicting its behavior, they certainly anthropomorphize the robot. However, anthropomorphism seems to be a broader concept involving attribution of various human traits, while adopting the intentional stance refers more narrowly to adopting a strategy in predicting and explaining others' behavior with reference to mental states.

INTENTIONAL STANCE AS A CRUCIAL FACTOR IN ACTIVATION OF MECHANISMS OF SOCIAL COGNITION IN HRI?

Currently, research in the fields of social robotics and HRI is exploring the effect of attributing intentionality to robots, and the behavioral parameters of the robot that most efficiently induce this. In exploring the effect that attributing intentionality has on social interaction, participants of an experiment are sometimes led to believe that they are interacting either with a pre-programmed machine (e.g., Wykowska et al., 2014; Özdem et al., 2017) or with another human (who naturally has desires and beliefs). In some other studies, participants are first exposed to different agent types (e.g., human, humanoid robot, non-humanoid robot), and subsequently are led to believe that they are interacting with one of them (e.g., Krach et al., 2008). Inducing a particular (e.g., intentional) stance through instruction manipulation stands in contrast to methods used in research that aims to define the parameters under which participants spontaneously assume the intentional stance. Here, the intentional stance is induced through, for example, the robot's gaze, speech, or general behavior (e.g., Wykowska et al., 2015).

Manipulating Adoption of Intentional Stance Through Explicit Instruction

Attitude or stance toward an artificial agent can be a major factor influencing the effectiveness of the subsequent interaction (e.g., Chaminade et al., 2012). Depending on beliefs concerning the processes underlying the observed behavior, people will use different strategies in interacting with them, and different brain regions will be activated during this interaction. An illustration of this comes from studies using economic games, in which participants were informed that they were playing a game together either with a computer or with another person (Kircher et al., 2009; Sripada et al., 2009). Here, participants use different strategies when playing the economic games; e.g., cooperating less readily in a prisoner's dilemma game and not engaging in punishing behavior in an ultimatum game when interacting with a machine (Rilling et al., 2004). On a neural level, brain regions associated with social cognition showed increased activation when participants believed they were engaging with a human as opposed to a machine (McCabe et al., 2001; Gallagher et al., 2002; Rilling et al., 2004). These results suggest that social cognitive mechanisms are more readily engaged in solving multi-agent problems, if the opponent is believed to be a human (see Schurz et al., 2014 for a meta-analysis). Although not all of these papers use Dennett's terminology, the belief that one's interaction partner is a human implies adoption of the intentional stance.

This trend extends beyond relatively high-level processes like decision making, since low-level psychological mechanisms such as sensory processing and attentional selection are similarly affected by whether one adopts the intentional stance or not (Wiese et al., 2012; Wykowska et al., 2014; Caruana et al., 2017). Studies have shown increases in neural signatures for sensory gain in response to social stimuli that are thought to be controlled by an intentional agent. Further, the effect of adopting the intentional stance on neuropsychological processes has been demonstrated in neuroimaging data. Brain regions related to mentalizing show more activation when participants interact with a robotic avatar that they believe is controlled by a person, than when they interact with this same avatar, but believe that its actions are preprogramed (Özdem et al., 2017). Activation of brain regions related to mentalizing have in turn been found to directly correlate with performance in tasks involving social cognition (Wiese et al., 2018).

These findings suggest that there are fundamental differences in how people interact with robots, based on the assumptions they hold concerning the cause of the robot's actions. Further, they demonstrate that interaction with non-human agents is positively influenced by the engagement of the brain's social capacities, which are increasingly activated when people believe that their interaction partner has a mind. Having adopted the intentional stance, humans will make use of the efficient information processing systems that social cognition offers, and social signals can be processed with increased efficiency. Social robots that are capable of inducing the intentional stance in their users might therefore be more capable of interacting with humans in a nigh-natural manner than those that cannot.

Inducing Adoption of Intentional Stance Through Interaction

In research which explores the direct effects of adopting the intentional stance on social interaction, intentional stance acts as an independent variable and is mostly induced through belief manipulation, while robot behavior is identical in the experimental conditions. Parallel to this, ongoing research examines the ways that the intentional stance can be induced in a more spontaneous manner as a consequence of robot behavior (Terada et al., 2008; Yamaji et al., 2010). When the means of inducing the intentional stance is itself the topic of research, a means of quantifying the success of this process becomes necessary, and although the intentional stance as a concept is defined at length (Dennett, 1971, 1987), measuring its adoption still presents a challenge. Additionally, much of the literature on the induction of the intentional stance comes from the engineering disciplines, and has a different approach to the earlier discussed research, which is based in experimental psychology. Despite the apparent similarity in overall goals and terminology, the methodology and research questions in these respective fields are often quite discrepant. A particularly prominent type of paradigm in HRI research into intentional attitudes involves naturalistic and relatively open-ended experiments (e.g., Terada et al., 2007; Yang et al., 2015). Participants in these studies are generally not given strict instructions to perform a certain task with the robot in question, rather leaving the interaction to develop naturally. Despite the similarity between this type of setup and the envisaged goal environment of a robotic platform, experimental validity is often compromised this way, and questions about what aspect of the robot's behavior lead to the adoption of the intentional stance remain unanswered.

Finally, despite discussing the circumstances under which the intentional stance is adopted and the consequent role this stance plays in moderating social cognitive mechanisms separately, these phenomena, of course, do not exist in isolation. This is illustrated by Pfeiffer et al. (2011), who demonstrated that estimates of humanness concerning an on-screen avatar in a social gaze experiment varied with a combination of gaze behavior and subject expectations concerning the avatar. Estimates of humanness increased when the gaze behavior of the avatar seemed to fit with the strategy that it was pursuing, according to the task instruction. Therefore, it is important to note that ultimately, it is a combination of behavioral parameters of, and a person's expectations toward the robot that together inform estimates of humanness. Presently, this is rarely taken into consideration in HRI research.

METHODS TO ASSESS ADOPTION OF INTENTIONAL STANCE

A major obstacle that affects all research into the parameters that induce adoption of the intentional stance is the lack of a reliable method to assess when people do in fact adopt the intentional stance. Specifically, to determine what behavioral parameters (e.g., Wykowska et al., 2015) induce intention attribution, it would be beneficial to develop a method that has the sensitivity to detect changes in attitudes resultant from robot behavior (i.e.,

in a pre-test vs. post-test manner). In this context, an interesting approach to measuring adoption of intentional stance has been reported by Thellman et al. (2017) or de Graaf and Malle (2018). In their paradigms, participants rated behaviors performed by either humans or robots in terms of, among other aspects, intentionality. The authors found that perceived intentionality of behaviors performed by robots, largely, closely matched that of identical behaviors performed by humans (Thellman et al., 2017; de Graaf and Malle, 2018). Note, however, that these findings come from surveys in which participants consider behaviors based on either written scenarios or pictures, and therefore reflect a conceptual consideration of the behavior, and not a direct reaction to it (Fussell et al., 2008).

Other methods have been used with varying levels of subtlety, few of which have undergone thorough validation. This lack of a common criterion makes comparison of studies on this topic especially troublesome, and precludes the use of common quantitative meta- analyses such as the computing of effect sizes (see Steinfeld et al., 2006; Weiss and Bartneck, 2015 for exceptions). Some researchers infer adoption of intentional stance indirectly by means of a different variable, which might be vulnerable to the influence of confounding variables, for instance, by the degree of successful interaction with a robot (Yamaji et al., 2010). Here, the authors found that children would often deposit trash into mobile trash cans when these trash cans moved toward litter on the ground and make twisting and bowing movements, signaling that this litter needed to be disposed of. Yamaji et al. (2010) concluded that this constituted a successful induction of intention attribution, as users were able to interpret what action was required from them. Alternatively, intention attribution has been inferred by the degree to which participants reported feeling "deceived" by a robot displaying unexpected behavior during an interactive game, which increased the likelihood that the robot would win (Terada and Ito, 2010). These authors argued that feeling deceived implies attributing intention to the robot.

The most intuitive method to measure the adoption of intentional stance involves simply asking participants to reflect on their interaction with a robot (Terada et al., 2008). Sirkin et al. (2015) deduced the degree of intentionality that participants attributed to the robot from a semi-structured interview. Terada et al. (2007) took a more straightforward approach, simply asking participants which of Dennett's stances they adopted in a brief interaction with a remote controlled chair. This method has its drawbacks, in that it requires introspection on the part of the participants as well as a moderate level of knowledge concerning Dennett's terminology. Introspection, though, is a notoriously unreliable research tool (Nisbett and Wilson, 1977; Fiala et al., 2014). Other studies make use of unstructured interviews, and subject the respondents' answers to analysis (Sirkin et al., 2015). Despite not requiring the deep level of introspection that is involved in asking participants to describe which of Dennett's stances they have adopted, this measure is still both explicit and reliant on self-reporting. Furthermore, being a qualitative technique, it does not lend itself well to comparative analysis between conditions or between studies.

Questionnaires are a common tool to obtain a quantitative measure of attitudes toward a robot, which avoids the pitfalls of
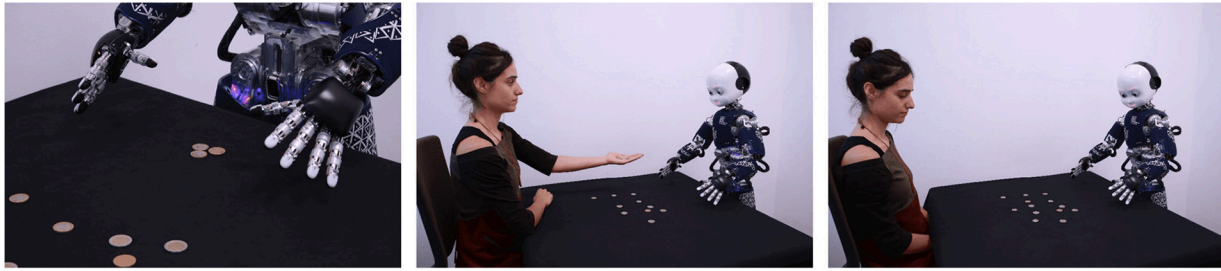
FIGURE 1 | Example scenario from the questionnaire of Marchesi et al. (2018).

the earlier described techniques. Questionnaires in HRI typically cover multiple aspects concerning the robot's perceived likability, utility, and humanness/anthropomorphism. Examples are the Psychological Humanoid Impression iTems (PHIT, Kamide et al., 2013), the Godspeed Questionnaire Series (GSQ, Bartneck et al., 2009) and the Robotic Social Attributes Scale (RoSAS, Carpinella et al., 2017). None of these are specifically aimed at assessing the adoption of the intentional stance toward a robot, though the PHIT has a subscale (Agency) that relates to the robot having agency and intentions of its own, although this subscale has not been explicitly validated as a corollary of adopting the intentional stance (Koda et al., 2016). A novel questionnaire has been designed specifically to assess the adoption of the intentional stance in HRI (Marchesi et al., 2018). This questionnaire consists of short series of photographs depicting the iCub robot (Metta et al., 2010) in various action sequences, see **Figure 1** for example scenario. The questionnaire provides two descriptions of the events in each series, one offering an explanation in terms of the mechanics and design of the robot, and the other in terms of intentions, beliefs and desires. Participants are asked to choose between the two alternative interpretations of the depicted scenario.

The story line presented in the scenarios usually does not offer a unique interpretation, and has an "open ending." In the example given, the description below the photographs reads:

"iCub receives three coins. iCub looks at the three coins.

1. Because iCub thinks it deserves more
2. Because iCub is counting."

The utterance was designed to be ambiguous in order to make each interpretation (mentalistic: option 1 in the above example, vs. mechanistic: option 2 in the above example) equally plausible. This questionnaire offers a dedicated research tool that does not necessitate introspection or knowledge about Dennett's taxonomy of stances on the part of the respondents, and is to some degree implicit, in that it does not directly ask participants to pass judgment on the robot's qualities. Still, this method has its drawbacks. The questionnaire's items depict one type of robot. This limits its generalizability, given the large degree of heterogeneity with which different robotic platforms are perceived (Mathur and Reichling, 2016). Additionally, it relies on conceptual judgments of a depicted behavior and does not reflect mentalizing processes that arise from a direct interaction with a robot.

Ideally, future research will be able to offer a still more implicit measure of intentional stance adoption. This may come in the form of paradigms that analyze reaction time and accuracy to assess underlying attitudes or stances, much like the implicit association test (Greenwald and Banaji, 1995; Luo et al., 2006). Additionally, research into neuropsychological markers and correlates of the intentional stance can explore the neural basis of assuming the intentional stance, offering ways to circumvent behavioral measures altogether (Spunt et al., 2015). Having a reliable technique to measure adoption of the intentional stance which can be compared across experiments, combined with a more methodical and controlled approach in studies attempting to parametrize robot behavior that gives rise to adoption of intentional stance, will help toward creating a comprehensive taxonomy of effective robot social behavior.

CONCLUSION

In sum, adoption of intentional attitudes in HRI is a topic to be addressed, as it might facilitate and enhance the effectiveness of social cognitive mechanisms, and thereby the overall quality and efficiency of human-robot communication. This has implications for robot design, as robots that are able to spontaneously induce intentional attitudes in their users will most likely be more effective communicators. Current research into robot behavioral parameters that induces this intentional attitude in users has certain limitations. For example, the lack of a reliable and common method to measure the adoption of intentional attitudes seriously impedes the validity of experimental findings and the comparability of different studies. Without a consensus on what constitutes adoption of the intentional stance and how one can quantify this, researchers in the field of HRI will be using different measuring rods. We propose that after having established well-validated methods for measuring adoption of the intentional stance, it would be of great importance to understand the conditions under which the intentional stance is adopted toward artificial agents.

AUTHOR CONTRIBUTIONS

ES performed literature research and review. AW conceived of the article's topic and structure, revised the manuscript and provided feedback with guidance on its writing.

FUNDING

REFERENCES

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81. doi: 10.1007/s12369-008-0001-3

Broadbent, E., Stafford, R., and MacDonald, B. (2009). Acceptance of healthcare robots for the older population: review and future directions. *Int. J. Soc. Robot.* 1:319. doi: 10.1007/s12369-009-0030-6

Cabibihan, J. J., Javed, H., Ang, M., and Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *Int. J. Soc. Robot.* 5, 593–618. doi: 10.1007/s12369-013-0202-2

Carpinella, C. M., Wyman, A. B., Perez, M. A., and Stroessner, S. J. (2017). "The robotic social attributes scale (rosas): development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 254–262.

Caruana, N., de Lissa, P., and McArthur, G. (2017). Beliefs about human agency influence the neural processing of gaze during joint attention. *Soc. Neurosci.* 12, 194–206. doi: 10.1080/17470919.2016.1160953

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutscher, E., Cheng, G., et al. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Front. Hum. Neurosci.* 6:103. doi: 10.3389/fnhum.2012.00103

de Graaf, M., and Malle, B. F. (2018). "People's judgments of human and robot behaviors: a robust set of behaviors and some discrepancies," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL: ACM), 97–98.

Dennett, D. C. (1971). Intentional systems. *J. Philos.* 68, 87–106. doi: 10.2307/2025382

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

Fiala, B., Arico, A., and Nichols, S. (2014). "You, robot," in *Current Controversies in Experimental Philosophy*, eds E. Machery and E. O'Neill (New York, NY: Routledge), 31–47.

Fussell, S. R., Kiesler, S., Setlock, L. D., and Yew, V. (2008). "How people anthropomorphize robots," in *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on IEEE* (Amsterdam: IEEE), 145–152.

Gallagher, H. L., Jack, A. I., Roepstorff, A., and Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage* 16, 814–821. doi: 10.1006/nimg.2002.1117

Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102:4. doi: 10.1037/0033-295X.102.1.4

Kamide, H., Kawabe, K., Shigemi, S., and Arai, T. (2013). Development of a psychological scale for general impressions of humanoid. *Adv. Robot.* 27, 3–17. doi: 10.1080/01691864.2013.751159

Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., et al. (2009). Online mentalising investigated with functional MRI. *Neurosci. Lett.* 454, 176–181. doi: 10.1016/j.neulet.2009.03.026

Koda, T., Nishimura, Y., and Nishijima, T. (2016). "How robot's animacy affects human tolerance for their malfunctions," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch: IEEE Press), 455–456.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3:e2597. doi: 10.1371/journal.pone.0002597

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., and Blair, R. J. R. (2006). The neural basis of implicit moral attitude—an IAT study using event-related fMRI. *Neuroimage* 30, 1449–1457. doi: 10.1016/j.neuroimage.2005.11.005

Marchesi, S., Ghiglino, D., Ciardo, F., Baykara, E., and Wykowska, A. (2018). Do we adopt the Intentional Stance toward humanoid robots? *PsyArXiv [Preprint].* doi: 10.31234/osf.io/6smkq

Mathur, M. B., and Reichling, D. B. (2016). Navigating a social world with robot partners: a quantitative cartography of the Uncanny Valley. *Cognition* 146, 22–32. doi: 10.1016/j.cognition.2015.09.008

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. (2001). A functional imaging study of cooperation in two person reciprocal exchange. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11832–11835. doi: 10.1073/pnas.211415698

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The iCub humanoid robot: an open-systems platform for research in cognitive development. *Neural Networks* 23, 1125–1134. doi: 10.1016/j.neunet.2010.08.010

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84:231. doi: 10.1037/0033-295X.84.3.231

Özdem, C., Wiese, E., Wykowska, A., Müller, H., Brass, M., and Van Overwalle, F. (2017). Believing androids–fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Soc. Neurosci.* 12, 582–593. doi: 10.1080/17470919.2016.1207702

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., and Schilbach, L. (2011). A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS ONE* 6:e27591. doi: 10.1371/journal.pone.0027591

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703. doi: 10.1016/j.neuroimage.2004.04.015

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. doi: 10.1016/j.neubiorev.2014.01.009

Sirkin, D., Mok, B., Yang, S., and Ju, W. (2015). "Mechanical ottoman: how robotic furniture offers and withdraws support," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM), 11–18.

Spunt, R. P., Meyer, M. L., and Lieberman, M. D. (2015). The default mode of human brain function primes the intentional stance. *J. Cogn. Neurosci.* 27, 1116–1124. doi: 10.1162/jocn_a_00785

Sripada, C. S., Angstadt, M., Banks, S., Nathan, P. J., Liberzon, I., and Phan, K. L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport* 20:984. doi: 10.1097/WNR.0b013e32832d0a67

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al. (2006). "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, UT: ACM), 33–40.

Terada, K., and Ito, A. (2010). "Can a robot deceive humans?," in *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (Osaka: IEEE Press), 191–192.

Terada, K., Shamoto, T., and Ito, A. (2008). "Human goal attribution toward behavior of artifacts," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on IEEE* (Munich: IEEE), 160–165.

Terada, K., Shamoto, T., Ito, A., and Mei, H. (2007). "Reactive movements of non-humanoid robots cause intention attribution in humans," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on IEEE* (San Diego, CA: IEEE), 3715–3720.

Thellman, S., Silvervarg, A., and Ziemke, T. (2017). Folk-psychological interpretation of human vs. humanoid robot behavior: exploring the intentional stance toward robots. *Front. Psychol.* 8:1962. doi: 10.3389/fpsyg.2017.01962

Weiss, A., and Bartneck, C. (2015). "Meta analysis of the usage of the Godspeed Questionnaire Series," in *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on IEEE* (Kobe: IEEE), 381–388.

Wiese, E., Buzzell, G., Abubshait, A., and Beatty, P. (2018). Seeing minds in others: mind perception modulates social-cognitive performance and relates to ventromedial prefrontal structures. *PsyArXiv [Preprint].* doi: 10.31234/osf.io/ac47k

Wiese, E., Metta, G., and Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Front. Psychol.* 8:1663. doi: 10.3389/fpsyg.2017.01663

Wiese, E., Wykowska, A., Zwickel, J., and Müller, H. J. (2012). I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PLoS ONE* 7:e45391. doi: 10.1371/journal.pone.0045391

Wykowska, A., Chamiande, T., and Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philos. Transac. Roy. Soc. Lond. B Biol. Sci.* 371:20150375. doi: 10.1098/rstb.2015.0375

Wykowska, A., Kajopoulos, J., Obando-Leitón, M., Chauhan, S., Cabibihan, J. J., and Cheng, G. (2015). Humans are well tuned to detecting agents among non-agents: examining the sensitivity of human perception to behavioral characteristics of intentional agents. *Int. J. Soc. Robot.* 7, 767–781. doi: 10.1007/s12369-015-0299-6

Wykowska, A., Wiese, E., Prosser, A., and Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS ONE* 9:e94339. doi: 10.1371/journal.pone.0094339

Yamaji, Y., Miyake, T., Yoshiike, Y., De Silva, P. R., and Okada, M. (2010). "STB: human-dependent sociable trash box," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on IEEE* (Osaka: IEEE), 197–198.

Yang, S., Mok, B. K. J., Sirkin, D., Ive, H. P., Maheshwari, R., Fischer, K., et al. (2015). "Experiences developing socially acceptable interactions for a robotic trash barrel," in *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on IEEE* (Kobe: IEEE), 277–284.