frontiers
in Computer Science

OPINION
published: 05 July 2021
doi: 10.3389/fcomp.2021.671012

What Sort of Robots Do We Want to Interact With? Reflecting on the Human Side of Human-Artificial Intelligence Interaction

Elisabeth Hildt *

Illinois Institute of Technology, Chicago, IL, United States

Keywords: ethics, social robots, human-technology interaction, social cognition, terminology, capabilities, human-robot interaction, social implications

INTRODUCTION

During the past decades, the interplay between humans and robots has been investigated in the field of human-robot interaction (HRI). This research has provided fascinating results on the spectrum and forms of engagement between humans and robots and on the various behaviors displayed by robots aimed at interacting with and influencing humans (Tsarouchi et al., 2016; Ahmad et al., 2017; Saunderson and Nejat, 2019a). Yet, crucial questions regarding how humans want to interact with and be influenced by robots are sidestepped in this research, falling for what could be called a robotistic fallacy. This article outlines some of the current findings on HRI to then critically assess the broader implications of HRI and key questions that must be asked in this context.

Social robots, i.e., robots that engage on a social level with humans, are expected to increasingly assist and support humans in workplace environments, healthcare, entertainment, training and education, and other fields (Ahmad et al., 2017; Richert et al., 2018; Pepito et al., 2020).

By using an interdisciplinary approach that involves behavioral studies and cognitive and social neuroscience, recent research on social cognition and HRI investigates how humans perceive, interact with and react to robots in social contexts (Cross et al., 2019; Henschel et al., 2020). Especially in the context of possible future uses in healthcare or geriatric care, the importance of developing robots with which humans can easily and naturally interact has been stressed (Pepito et al., 2020; Wykowska 2020).

Henschel et al. (2020) argue that research into and knowledge of the neurocognitive mechanisms involved in human-robot interaction will supply critical insights for optimizing social interaction between humans and robots which will in turn help to develop socially sophisticated robots. They write (Henschel et al., 2020, p. 373): "Robots that respond to and trigger human emotions not only enable closer human-machine collaboration, but can also spur human users to develop long-term social bonds with these agents." This approach suggests using cognitive neuroscience to build robots that humans are likely to emotionally interact with, an approach that can be seen as an extension of affective computing (Scheutz, 2011; McDuff and Czerwinski, 2018). In this, the focus is on building social robots so that HRI resembles human-human interaction (HHI). A question rarely asked though, is how humans would like to interact with robots and what sort of robots humans would like to interact with.

For example, Wiese et al. (2017) argue that in order for humans to interact intuitively and socially with robots, robots need to be designed in a way that humans perceive them as intentional agents, i.e., as agents with mental states. They elaborate that this is achieved when robots evoke mechanisms of social cognition in the human brain that are typically evoked in HHI. Consequently, they advocate for integrating behavioral and physiological neuroscience methods in the design and evaluation of

OPEN ACCESS

Edited by:
Manuel Ruiz Galan,
University of Granada, Spain

Reviewed by:
Gabriel Skantze,
Royal Institute of Technology, Sweden

***Correspondence:**
Elisabeth Hildt
ehildt@iit.edu

Received: *22 February 2021*
Accepted: *21 June 2021*
Published: *05 July 2021*

Citation:
Hildt E (2021) What Sort of Robots Do We Want to Interact With? Reflecting on the Human Side of Human-Artificial Intelligence Interaction.
Front. Comput. Sci. 3:671012.
doi: 10.3389/fcomp.2021.671012

social robots to build robots that are perceived as social companions. A questionnaire-based study by Marchesi et al. (2019) found that at least sometimes, humans adopt the intentional stance to humanoid robots by explaining and predicting robot behavior through mental states and mentalistic terms. Ciardo et al. (2020) investigated participants' sense of agency, i.e., the perceived control felt on the outcome of an action, when participants engaged with robots. Their results indicate that in HRI involving a shared control setting, participants perceived a lower sense of agency, similar to what can be observed in comparable HHI.

BROADER IMPLICATIONS OF HUMAN-ROBOT INTERACTION

These are but a few examples of recent studies in an upcoming research field. Overall, research indicates that humans tend to interact with social robots in similar ways as they interact with humans, with anthropomorphic features facilitating this effect. From the perspective of HRI research, these similarities are considered an advantage, the goal being to build robots with which humans can easily and intuitively interact. In this, the guiding assumption is that cognitive and behavioral studies serve as the basis for building robots with which humans engage in such a way that HRI resembles HHI.

However, this way of reasoning jumps too easily from the empirical observation that HRI resembles HHI to the normative conclusion that HRI should resemble HHI. It could thus be called a robotistic fallacy. A possible reply to this criticism is that what makes HRI that resembles HHI attractive, is that it is user-friendly, allows for effective interaction with robots, and provides the basis for social acceptance of robots. However, so far, these claims are essentially unproven. Further, they rely on very narrow conceptions of user-friendliness and social interaction, as will be outlined below.

While the focus of most HRI studies has been on how HRI is similar to HHI (Irfan et al., 2018; Henschel et al., 2020), there is also the uncanny valley hypothesis, according to which robots with a too anthropomorphic design are considered disturbing (Mori et al., 2012; Richert et al., 2018). Furthermore, current research is primarily confined to laboratory situations and investigates the immediate situation of HRI in experimental settings. When considering real life situations, a multitude of additional factors will come in that have not been researched yet (Jung and Hinds, 2018). Interaction with robots "in the wild" will probably turn out to be much messier and more complex than research studies that highlight and welcome similarities between HRI and HHI currently assume. Reflections on the broader implications of HRI on humans go beyond the immediate experimental setting and include the broader social context. In this context, three aspects are worth considering:

Robot Capabilities

While an immediate HRI can resemble HHI, robots differ in crucial ways from humans. Current robots do not have capabilities that are in any way comparable to human sentience, human consciousness, or a human mind. Robots only *simulate* human behavior. Humans tend to react to this simulated behavior in similar ways as they react to human behavior. This is in line with research according to which humans interact with computers and new media in fundamentally social and natural ways (Reeves and Nass, 2002; Guzman, 2020).

However, capabilities matter. While taking the intentional stance toward robots may help to explain robot behavior and facilitate an interaction with robots, it does not say much about robot capabilities or the quality of the interaction. Superficially, in certain situations, the reactions of a robot simulating human behavior and human emotions and a person having emotions and showing a certain behavior may be similar. But there is a substantial difference in that there is no interpersonal interaction or interpersonal communication with a robot. While this may not play a huge role in confined experimental settings, the situation will change with wider applications of social robots. Questions to be addressed include: What are the consequences of inadequate ascription of emotions, agency, accountability and responsibility to robots? How should one deal with unilateral emotional involvement, lack of human touch and absence of equal level interaction? And how may HRI that simulates HHI influence interpersonal interactions and relationships?

How to Talk About Robot Behavior?

While it may seem tempting to describe robot behavior that simulates human behavior with the same terms as human behavior, the terminology used when talking about robots and HRI clearly needs some scrutiny (see also Salles et al., 2020). For example, in HRI studies in which participants were asked to damage or destroy robots, robots were characterized as being "abused," "mistreated" or "killed" (Bartneck and Hu, 2008; Ackerman, 2020; Bartneck and Keijsers, 2020; Connolly et al., 2020). It is questionable, however, whether concepts like "abuse" or "death" can meaningfully be used for robots. The same holds for ascribing emotions to robots and talking of robots as "being sad" or "liking" something. Claims like "Poor Cozmo. Simulated feelings are still feelings!" (Ackerman, 2020) are clearly misleading, even if meant to express some irony.

In part, this problematic language use results from a strong tendency of anthropomorphizing that is directly implied by an approach that focuses on HRI resembling HHI. In part, it may be considered a use of metaphors, comparable to metaphorical descriptions in other contexts (Lakoff and Johnson, 2003). In part, issues around terminology may be a matter of definition. Depending on the definition given, for example for "mind" or "consciousness", claims that current robots do have minds or consciousness may be perfectly adequate, implying that robot mind or robot consciousness are significantly different from human-like mind or consciousness (Bryson, 2018; Hildt, 2019; Nyholm, 2020). It may be argued that as long as clear definitions are provided, and different conceptions are used for humans and robots, this type of language use is not problematic. However, it will be important to establish a terminology for robots that allows the use of concepts in such a way that there is no interference with the way these concepts are used for humans. Otherwise, the same term is used for two different things. As a result, humans may expect from robots that are characterized as "being conscious" or "having a mind" much more than is exhibited by the technology.

While these are primarily theoretical considerations for now, empirical studies will certainly find out more about the language used to talk about social robots and the implications.

Robots Influencing Humans

Reflections on the broader implication of HRI beyond laboratory settings include the question of what roles humans, individually and as a society, want to ascribe to robots in their lives, and how much they want to be influenced by robots. For example, in a recent exploratory HRI study (Saunderson and Nejat, 2019b), social robots used different behavior strategies in an attempt to persuade humans in a game, in which the human participants were asked to guess the number of jelly beans. The robots exhibited various verbal and nonverbal behaviors for different persuasive strategies, including verbal cues such as "It would make me happy if you used my guess (. . .)" to express affect, and "You would be an idiot if you didn't take my guess (. . .)" to criticize.

While this certainly is an interesting study, a number of questions come to mind: Is it realistic to assume that one can make a robot happy for taking its guess? In how far can a person be meaningfully blamed by a robot? What would it mean, upon reflection, to be persuaded by a robot? In how far is there deception involved? For sure, it is not the robot itself but the people who design, build and deploy the technology who attempt to elicit a certain human behavior. And there clearly are similarities to commercials and various forms of nudging. In settings like these, aspects to consider include the type of influence, its initiator, the intentions behind and the consequences of the interaction.

CONCLUSION

HRI research has shown that in various regards, humans tend to react to robots in similar ways as they react to human beings. While these are fascinating results, not much consideration has been given to the broader consequences of humans interacting with robots in real-life settings and the social acceptance of social robots. When it comes to potential future applications beyond experimental settings, a broader perspective on HRI is needed that better takes the social and ethical implications of the technology into account. As outlined above, aspects to be considered include how to deal with simulated human-like behavior that is not grounded in human-like capabilities and how to develop an adequate terminology for robot behavior. Most crucially, the questions of what social roles to ascribe to robots and to what extent influence exerted by robots would be considered acceptable need to be addressed. Instead of planning to build robots with which humans cannot but interact in certain ways, it is crucial to think about how humans would like to interact with robots. At the center of all of this is the question "What sort of robots do we want to engage with?" For it is humans who design, build and deploy robots and who by designing, building and deploying robots shape the ways humans interact with robots.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Ackerman, E. (2020). *Can Robots Keep Humans from Abusing Other Robots?* IEEE Spectrum. Available at: https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/can-robots-keep-humans-from-abusing-other-robots (Accessed August 19, 2020).

Ahmad, M., Mubin, O., and Orlando, J. (2017). A Systematic Review of Adaptivity in Human-Robot Interaction. *Mti* 1 (3), 14. doi:10.3390/mti1030014

Bartneck, C., and Hu, J. (2008). Exploring the Abuse of Robots. *Is* 9 (3), 415–433. doi:10.1075/is.9.3.04bar

Bartneck, C., and Keijsers, M. (2020). The Morality of Abusing a Robot. *J. Behav. Robotics* 11, 271–283. doi:10.1515/pjbr-2020-0017

Bryson, J. J. (2018). Patiency Is Not a Virtue: the Design of Intelligent Systems and Systems of Ethics. *Ethics Inf. Technol.* 20, 15–26. doi:10.1007/s10676-018-9448-6

Ciardo, F., Beyer, F., De Tommaso, D., and Wykowska, A. (2020). Attribution of Intentional agency towards Robots Reduces One's Own Sense of agency. *Cognition* 194, 104109. doi:10.1016/j.cognition.2019.104109

Connolly, J., Mocz, V., Salomons, N., Valdez, J., Tsoi, N., Scassellati, B., et al. (2020). "Prompting Prosocial Human Interventions in Response to Robot Mistreatment," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20); March 23–26, 2020, Cambridge, United Kingdom. Editors T. Belpaeme and J. Young (New York, NY, USA: ACM), 10. doi:10.1145/3319502.3374781

Cross, E. S., Hortensius, R., and Wykowska, A. (2019). From Social Brains to Social Robots: Applying Neurocognitive Insights to Human-Robot Interaction. *Phil. Trans. R. Soc. B* 374, 20180024. doi:10.1098/rstb.2018.0024

Guzman, A. (2020). Ontological Boundaries between Humans and Computers and the Implications for Human-Machine Communication. *Hmc* 1, 37–54. doi:10.30658/hmc.1.3

Henschel, A., Hortensius, R., and Cross, E. S. (2020). Social Cognition in the Age of Human-Robot Interaction. *Trends Neurosci.* 43 (6), 373–384. doi:10.1016/j.tins.2020.03.013

Hildt, E. (2019). Artificial Intelligence: Does Consciousness Matter? *Front. Psychol.* 10, 1535. doi:10.3389/fpsyg.2019.01535

Irfan, B., Kennedy, J., Lemaignan, S., Papadopoulos, F., Senft, E., and Belpaeme, T. (2018). "Social Psychology and Human-Robot Interaction: An Uneasy Marriage," in HRI '18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion; Chicago, IL, USA, March. 5–8, 2018. New York, NY, USA: ACM, 8. doi:10.1145/3173386.3173389

Jung, M., and Hinds, P. (2018). Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *ACM Trans. Hum.-Robot Interact.* 7 (1), 5. doi:10.1145/3208975

Lakoff, G., and Johnson, M. (2003). *Metaphors We Live by*. Chicago and London: University of Chicago Press. doi:10.7208/chicago/9780226470993.001.0001

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., and Wykowska, A. (2019). Do We Adopt the Intentional Stance toward Humanoid Robots? *Front. Psychol.* 10, 450. doi:10.3389/fpsyg.2019.00450

McDuff, D., and Czerwinski, M. (2018). Designing Emotionally Sentient Agents. *Commun. ACM* 61 (12), 74–83. doi:10.1145/3186591

Mori, M., MacDorman, K., and Kageki, N. (2012). The Uncanny valley [from the Field]. *IEEE Robot. Automat. Mag.* 19 (2), 98–100. doi:10.1109/mra.2012.2192811

Nyholm, S. (2020). *Humans and Robots. Ethics, Agency, and Anthropomorphism*. London, New York: Rowman & Littlefield.

Pepito, J. A., Ito, H., Betriana, F., Tanioka, T., and Locsin, R. C. (2020). Intelligent Humanoid Robots Expressing Artificial Humanlike Empathy in Nursing Situations. *Nurs. Philos.* 21 (4), e12318. doi:10.1111/nup.12318

Reeves, B., and Nass, C. (2002). *The Media Equation.How People Treat Computers, Television, and New Media like Real People and Places*. Stanford: CSLI Publications.

Richert, A., Müller, S., Schröder, S., and Jeschke, S. (2018). Anthropomorphism in Social Robotics: Empirical Results on Human-Robot Interaction in Hybrid Production Workplaces. *AI Soc.* 33, 413–424. doi:10.1007/s00146-017-0756-x

Salles, A., Evers, K., and Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neurosci.* 11 (2), 88–95. doi:10.1080/21507740.2020.1740350

Saunderson, S., and Nejat, G. (2019a). How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human-Robot Interaction. *Int. J. Soc. Rob.* 11, 575–608. doi:10.1007/s12369-019-00523-0

Saunderson, S., and Nejat, G. (2019b). It Would Make Me Happy if You Used My Guess: Comparing Robot Persuasive Strategies in Social Human-Robot Interaction. *IEEE Robot. Autom. Lett.* 4 (2), 1707–1714. doi:10.1109/lra.2019.2897143

Scheutz, M. (2011). "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots," in *Robot Ethics. The Ethical and Social Implications of Robotics*. Editors P. Lin, K. Abney, and G. A. Bekey (USA: MIT Press), 205–221.

Tsarouchi, P., Makris, S., and Chryssolouris, G. (2016). Human-robot Interaction Review and Challenges on Task Planning and Programming. *Int. J. Computer Integrated Manufacturing* 29 (8), 916–931. doi:10.1080/0951192x.2015.1130251

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

Wykowska, A. (2020). Where I Work. *Nature* 583, 652. https://media.nature.com/original/magazine-assets/d41586-020-02155-1/d41586-020-02155-1.pdf

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.