Minds and Machines (2021) 31:329–336
https://doi.org/10.1007/s11023-021-09572-9

EDITORIAL

The Dawn of Social Robots: Anthropological and Ethical Issues

Georg Gasser[1]

Published online: 16 September 2021

1 Social Robots

For some time now robots have left the confines of industry to work more closely together with humans: "cobotics" is now a common buzzword. Robots of a new technological generation are currently on the threshold of everyday, practical use in the service, health and education sectors. The International Federation of Robots (IFR) predicts that autonomous mobile robots (AMRs) in the logistics sector, for example, will anually increase by 30% between 2020 and 2023.[1]

As robots become more mobile thanks to sensors, actuators and artificial intelligence, enabling them to move in a more complex environment, direct interaction with people in an everyday setting becomes increasingly normal. As their autonomy increases, they can independently work through hierarchies of goals and are less under ongoing human control and intervention. For example, autonomous robots have been assisting in shopping malls in Japan and Korea for some time now; in Korea, thousands of such robots work as teaching assistants in kindergartens (Han et al., 2015), and the Henn-na Hotel in Nagasaki has become the first hotel to rely extensively on robots to perform the tasks normally performed by hotel employees such as welcoming guests at the reception, providing information, organising the storage of luggage and taking care of the check-in and check-out services.

As part of this "mobile robot revolution", social robots have established themselves as a distinct group. Conceptually, these are robots that are specifically equipped with social-communicative functions, since their primary task is to react to humans as social and communicative beings and to refer particularly to these essential facets of human existence. Social robots should ideally recognize the social needs of humans together with the associated feelings and thoughts and respond appropriately according to their respective roles.

[1] https://ifr.org/ifr-press-releases/news/mobile-robots-revolutionize-industry (from August 5, 2021).

✉ Georg Gasser
georg.gasser@kthf.uni-augsburg.de

[1] Faculty of Catholic Theology & Department of Philosophy, University of Augsburg, Augsburg, Germany

The implementation of this aim seems particularly urgent because the current demographic trend in industrialised countries means that the number of older people who will be dependent on care in one form or another will rise rapidly in the coming years. At the same time, however, the number of available caregivers is constantly decreasing, which is why social robots are seen as one way to respond to this worrisome development. In addition to robots that cover the area of medical care (e.g. monitoring blood pressure, detecting falls, encouraging physical exercise...) and everyday care and hygiene (washing, bathing...), it is also crucial that the cognitive-social needs of the persons to be cared for are not neglected (playing games, memory training, providing entertainment and companionship...).

The successful performance of such tasks does not necessarily have to be accompanied by a humanoid form. A washing robot for the care sector can also be equipped with communicative functions and thus also react to social needs in the context of personal hygiene. The social robots most frequently encountered in practice, such as Nao and Pepper, can clearly be classified as non-human in their appearance.

Previous experience with these robots suggests that appropriate interactions may well promote human communication and social skills. For example, AIBO, a touch-sensitive and interacting pet introduced by Sony more than two decades ago, has been successfully tested as an activating toy for the elderly. Similarly, Paro, a Japanese companion robot shaped like a baby harp seal, assists elderly people and patients. Paro responds to petting with body movements as well as by opening and closing its eyes and making sounds. In general, people like to cuddle, stroke, and talk with Paro like they would with a real animal. The use of Paro is analogous to real therapy animals, of which empirical studies show that interaction with them can help to lower blood pressure, cushion depressive phases, reduce the subjective feeling of pain, improve the phase of recovery or bring someone out of social self-isolation. For example, there is also evidence that Paro can reduce feelings of loneliness and social isolation in retirement houses or alleviate emotional agitation, stress and depressive phases in psychogeriatric patients and patients suffering from dementia (Jøranson et al., 2015; Góngora Alonso et al., 2019). Thus, in the fields of health- and social care, such robots can be seen as a useful tool: They improve existing solutions or allow therapists and caregivers new areas of application.

As social robots come closer to their users than previous machines due to their communicative and social abilities in the personal sphere, a whole range of ethical questions arise: Can social robots be a social counterpart in a genuine way, perhaps even take on the role of a human friend? Is it better to have a robot as "Ersatz-partner" than to have to live in loneliness? If the artificial seal Paro has a positive influence on elderly people, do they have to be informed (if this is possible at all) that it is an artificial and not a real animal? How are personal rights protected when interaction with a robot continuously collects data to adapt and improve the interaction? Are we in danger of increasingly delegating interpersonal contacts in nursing homes to robots? Are there interpersonal areas that should be off-limits to robots? Is there such a thing as respect that is appropriate towards robots or do they even enjoy a certain moral status?

In my introductory reflections, I do not want to go further into these or similar questions. They would go beyond the scope of this introduction and, in addition,

some of them will be discussed in detail in the papers that follow. Therefore, I would like to focus on one crucial anthropological issue, which concerns our embodiment: We are essentially embodied beings and the structure of our body shapes deeply our mind, that is, how we perceive ourselves, others and the world. This fact also affects our perception of robots, especially those designed in their appearance and behaviours as human-like as possible.

2 Embodiment and Social Cognition

The thesis that consciousness is intimately linked to our biological constitution has become a commonplace of contemporary Philosophy of Mind. The so-called Philosophy of Embodiment can be understood as a broad research programme that revolves around the central thesis that both the cognitive and mental states and processes of living beings are intrinsically embodied and as such essentially embedded in an environment (e. g. Haugeland, 1998; Varela et al., 1991).

This claim corresponds to the phenomenological insight that our body is not simply an instrument controlled and moved by the mind, similar to a captain steering an airplane. Rather, the concrete nature of the body and its being embedded in a surrounding environment essentially shapes our mental life. A body is not simply a body with which I am connected in a causal-external way, but I experience the body directly as something subjectively accessible, which is present in perceiving, feeling, thinking and acting and which characterizes the way I am "in the world". I may regard my body as "a thing among things" but nevertheless my body as a body occupies a unique position vis-à-vis me in that I cannot distance myself from it as I do from other things (Husserl, 1973, p. 162). I can put other things away, while my body is always with me. I experience my body as a living body (*Leib*), that is, "as something directly alive and connected to me, and as such it has a fundamentally different kind of experience for me than all other bodies have for me" (Husserl, 2008, p. 615).

We do neither encounter minds that are not embodied nor agents that are not interacting with the world through their bodies. We are, as Alva Noe puts it, not in our head but interacting with our environment as "distributed, dynamically spread-out, world-involving beings" (Noe, 2009, p. 9).

These considerations have direct relevance to the way we understand social cognition and interpersonal interaction. Traditional accounts of theory of mind, for instance, have it that there is a gap between the mental life of two persons and this gap is bridged by some kind of cognitive processes in one mind providing the means to infer the activities going on in the other mind. Thus, the mind of another person is not directly perceptible but concealed by the body. What one needs to bridge this gap is either a kind of theory about other minds or a kind of simulation of other minds or a combination thereof that will permit an inferential form of mind-reading or mind-simulating.

Without going into the details of these approaches, it should be obvious that the body plays a subordinate role here. The role it has, at best, is to be the source of evidence for constructing the relevant inference. Social cognition is characterised as a

third-person process where one person is observing the behaviour of another person and then drawing corresponding inferences about her inner mental life hidden away.

The direct social cognition account, instead, argues that social cognition and interpersonal interaction takes places directly and not via inferences because the mind is usually something directly accessible. The idea is that if body and mind are intimately interconnected, then the body of another person provides us direct access to his or her mind: The posture, movements, face, voice, gestures or skin tones are expressions of a person's mind. Ludwig Wittgenstein highlights this point when he writes:

> Look into someone else's face, and see the consciousness in it, and a particular shade of consciousness. You see on it, in it, joy, indifference, interest, excitement, torpor, and so on. … Do you look into yourself in order to recognise the fury in *his* face? (Wittgenstein, 1967, p. 229)

This is not to say that we never draw on the aforementioned models in our social interactions, nor that we are always able to correctly capture the mental states of another person. The claim is rather that in most of our encounters in everyday life, direct perception delivers a significant amount of important information for understanding others and for being able to interact successfully with others.

This approach receives additional support from studies in developmental psychology, which indicate that already infants automatically attune to facial expression and voices with a mimetic response (Schilbach et al., 2008). Even before the first year of life, children are able to perceive various body movements as meaningful and goal-directed (Senju et al., 2006). These capacities do not require advanced cognitive abilities such as making inferences or simulations, which infants at that young age simply lack; rather, they are perceptual capacities that run automatically and are highly stimulus-driven. Without wanting to go into further developmental steps, it should only be noted at this point that these abilities are not lost in later development, but are supplemented by other higher-level cognitive abilities (a good overview provides Gallagher, 2008).

In short, we are in a position to interact with and to understand others in terms of their (contextualised) bodily expressions, gestures, vocal intonations and movements long before we are able to theorise, simulate, reflect upon or predict the mental states of others. Our interpersonal interaction is essentially perception-based because our minds are embodied and therefore, thanks to the body, the mental life of others becomes accessible to us.

3 Humanoid Robots and Social Interaction

Already 50 years ago a positive relationship between humanoid robots and feelings of comfort with them was proposed. Findings suggested, however, a steep dip in comfort when robots looked almost but not perfectly human, for instance, because of an odd way of moving (Mori, 1970). This dip is called the uncanny valley and it highlights negative feelings of unease, eeriness or even hostility towards human-like robots which are difficult to distinguish from real humans.

In literature, this phenomenon has already been described by the German writer E.T.A. Hoffmann (1776–1822) in the short story "The Automata", when one of the two protagonists, Lewis, is confronted with the talking Turk, an automaton with a very human appearance, and says: "For me, the very association of a human person with dead figures that imitate humanity in their formation and movement has something oppressive, uncanny, even horrifying about it. I can imagine that it would be possible to make figures dance artificially and nimbly by means of a mechanism hidden inside, and that they would have to perform a dance together with human beings, turning and twisting in all kinds of ways, so that the living dancer would take hold of the dead wooden dancer and swing with her. Would you be able to bear the sight for a minute without inner horror?" (my translation from the German original)

Precise information on the extent of the uncanny valley effect varies, as do the particular (outer) appearance and (inner) functioning of a robot that lead to an increase or decrease in the required humanlikness (e. g. Müller et al., 2021; Rosenthal-von der Pütten & Krämer, 2014; Zhang et al., 2020).

The "category uncertainty hypothesis" suggests that an important cause for unease towards human-like robots could be caused by category uncertainty, that is, the uncertainty of whether it is a human being or not. Conflicting cues that make a clear categorisation difficult may cause us to feel uncomfortable and therefore adopt negative reactions towards the problematic object (Wang et al., 2015). The "mind perception hypothesis" can be seen as a variant of the "category uncertainty hypothesis" because it proposes that humanoid robots are uncanny because they are so realistic that we tend to ascribe to these robots mental capacities such as feelings and thoughts although we are convinced that these capacities are unique characteristics of complex animals (Gray & Wegner, 2012).

Regardless of the concrete differentiation of these hypotheses, both point out that a no longer given clear distinction between humans on the one hand and robots on the other undermines our self-image and identity as human person, which results in the indicated negative reactions.

The brief remarks on embodiment support this view, since due to the humanoid appearance of the robot ("the embodiment of the robot") those stimuli are given that automatically activate those systems in us that are relevant for social interaction processes. However, at the same time we perceive other features in the robot that speak against the activation of these processes or we are even aware that we are dealing with an entity lacking any form of mental life. There are also corresponding findings from developmental psychology that suggest that infants react with stress or fear to objects such as hoover robots that can move on their own but do not have feet or other locomotion organs, since the fundamental distinction between living beings that can move on their own and non-living objects that have to be moved from the outside is suspended. In these cases, it is assumed that infants have a corresponding basic categorisation of self-moving living and non-self-moving non-living entities, which is not based upon theory-driven ascription, since the corresponding mental prerequisites for doing so do not yet exist at this young age.

In addition, humans have a tendency to see the world through an anthropomorphic filter. When trying to interact with a new and unfamiliar entity, human use the knowledge about themselves as a basis for interpreting and predicting the behaviour

of this entity. This tendency increases with the humanlikness of an entity and is further nurtured though our social needs. Evidence suggests that our social nature leads us to anthropomorphize humanoid objects more in situations of social loneliness than in situations where our social needs are largely met (Eyseel & Reich, 2013).

If these assumptions are correct, then, at least in the health and care sectors, it might be advisable to design social robots in such a way that they can be clearly distinguished from humans (there may be other areas of application where the best possible approximation to humans is indicated, but I am leaving this question aside here). Ultimately, the appearance of a robot also has a specific purpose to fulfil and in this area of application it should not be about replacing human carers or making the people being cared for believe that they are interacting with real human people. Studies suggest that the appearance of a robot should be appropriate to the tasks it has to perform: For example, animal-like, fluffy robots are preferred when addressing social needs and emotions, while more machine-like robots are appealing for administering medication, lifting or helping with washing (Broadbent et al., 2012).

Moreover, also non-humanoid robots can accommodate our embodied and agentive nature as they encourage interaction with them. As part of our environment, social robots offer opportunities for interaction and touch as basic channels to transmit feelings, affects and needs that are only available to a reduced extent when robots are not physically but only digitally present as avatars or via a screen. Humans interact more with a physically present robot than with digital representations of it, as this obviously corresponds to our embodied and social nature. For instance, studies indicate that people experience their interactions more engaging and effective and are also more compliant to follow instructions when the social robot is present than when they merely have to interact with the same software and voice on a screen (Mann et al., 2015; Deng et al., 2019).

4 Conclusions

Predictions by experts indicate that robots will become a natural part of our environment in the coming decades. Robots are already proving useful in taking over important tasks in the healthcare and social sector, thereby relieving human employees.

Engineers often strive to make robots look as human-like as possible so that we break down barriers to interaction and feel comfortable in their presence. However, the uncanny valley effect indicates that the presence of humanoid robots can also evoke feelings of discomfort, eeriness and threat. I have pointed out that our social nature plays an important role in this context. As socially oriented creatures, we are fundamentally interested in interaction and cooperation, which requires the reliable ability to understand the other side to a certain degree. Robots that are too humanoid are likely to prevent us from such an understanding or they make us feel insecure, which has a negative effect on our willingness to interact.

Our social nature is shaped in a significant way by our embodied constitution. Our embodiment determines the way we see and interpret ourselves and the world as well as the range of interaction possibilities we are given with the world and with others.

Robots principally represent additional possibilities for interaction, and if these robots can be clearly distinguished from humans, we do not experience this interaction in most situations as negative. We prefer physically present robots to comparable software counterparts because of our agentive and social orientation. Non-humanoid robots also indicate that robots are not meant to replace human interaction partners, but rather represent an additional complementary role in the field of social interaction and can support and promote it. Thus, while it is seen as positive that robots are able to demonstrate a certain degree of functional autonomy or express verbally understanding for the human counter-part or explain why they behave in a given way, moderation in the endowment of all-too-human attributes is key if one wishes to avoid negative reactions.

Ultimately, the central questions in the context of the further development of social robots and our relationships with them are of a philosophical nature: Does blurring the line between humans and robots challenge our personal identity? What are our purposes in designing robots that are as human-like as possible? Do we perhaps see in such robots a more ideal form of our own nature? If so, what does this tell us about ourselves and our relationships with others? These and similar questions address the articles included in this special issue of Minds and Machines.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Broadbent, E., Tamagawa, R., Patience, A., Knock, B., Kerse, N., et al. (2012). Attitudes towards healthcare robots in a retirement village. *Australasian Journal on Ageing, 31*, 115–120. https://doi.org/10.1111/j.1741-6612.2011.00551.x Epub 2011 Jul 21 PMID: 22676171.

Deng, E., Mutlu, B., & Mataric, M. J. (2019). Embodiment in Socially Interactive Robots. *Foundations and Trends® in Robotics, 7*(4), 51–356. https://doi.org/10.1561/2300000056

Eyssel, F., & Reich, N. (2013). Loneliness makes the heart grow fonder (of robots): on the effects of loneliness on psychological anthropomorphism. *Proceedings of the ACM/IEEE International Conference of Human-Robot Interact. 8th* (Tokyo, March 3–6), 121–22. IEEE.

Gallagher, S. (2008). Direct perception in the social context. *Consciousness and Cognition, 17*, 535–543.

Góngora Alonso, S., Hamrioui, S., de la Torre, D. I., Motta Cruz, E., López-Coronado, M., & Franco, M. (2019). Social robots for people with aging and dementia: A systematic review of literature. *Telemedicine Journal and E-Health, 25*(7), 533–540.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Han, J., Jo, M., Hyun, E., et al. (2015). Examining young children's perception toward augmented reality-infused dramatic play. *Educational Technology Research and Development, 63*, 455–474. https://doi.org/10.1007/s11423-015-9374-9

Haugeland, J. (1998). Mind embodied and embedded. In J. Haugeland (Ed.), *Having thought: essays in the metaphysics of mind* (pp. 207–237). Harvard University Press.

Husserl, E. (1973). *Ding und Raum. Vorlesungen 1907*. Edited by Ulrich Claesges (Husserliana 16). Springer.

Husserl, E. (2008). *Die Lebenswelt. Auslegungen der vorgegebenen Welt und ihrer Konstitution. Texte aus dem Nachlass (1916–1937)*. Edited by Rochus Sowa (Husserliana 39). Springer.

Jøranson, N., Pedersen, I., Rokstad, A. M., & Ihlebæk, C. (2015). Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: A cluster-randomized controlled trial. *Journal of the American Medical Directors Association, 16*, 867–873.

Mann, J. A., MacDonald, B. A., Kuo, I., Li, X., & Broadbent, E. (2015). People respond better to robots than computer tablets delivering healthcare instructions. *Computers in Human Behavior*, *43*, 112–17. https://doi.org/10.1016/j.chb.2014.10.029.

Mori, M. (1970). The uncanny valley. *Energy, 7*, 33–35.

Müller, B. C. N., Gao, X., Nijssen, S. R. R., et al. (2021). I, Robot: How human appearance and mind attribution relate to the perceived danger of robots. *Int J of Soc Robotics, 13*, 691–701. https://doi.org/10.1007/s12369-020-00663-8

Noe, A. (2009). *Out of Our Heads. Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Hill & Wang.

Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior, 36*, 422–439.

Schilbach, L., Eickhoff, S. B., Mojzisch, A., & Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience, 3*(1), 37–50.

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience, 1*(3–4), 220–234.

Varela, F. J., Thompson, E. T., & Rosch, E. (1991). *The embodied mind. Cognitive science and human experience*. MIT Press.

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Wittgenstein, L. (1967). *Zettel*. Edited by G. E. M. Anscombe & G. H. von Wright, trans. G. E. M. Anscombe. University of California Press.

Zhang J., et al, 2020: A Literature Review of the Research on the Uncanny Valley. In: Rau PL. (eds) *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments*. HCII 2020. Lecture Notes in Computer Science, vol 12192. Springer, 255–268. https://doi.org/10.1007/978-3-030-49788-0_19.