Deception and Manipulation in Social Robotics

Karolina Zawieska
Industrial Research Institute for Automation and Measurements PIAP
& SMARTlab/School of Education, University College Dublin (UCD)
Al. Jerozolimskie 202, 02-489 Warsaw, Poland
kzawieska@piap.pl

ABSTRACT
The core of anthropomorphism is illusion, and hence, one of the key ethical concerns in social robotics refers to deception. A common way to address this issue is to study the degree to which it is possible to influence humans through anthropomorphic robot design and whether using deception provides benefits to robot users. This work argues that the main ethical issue lies not in deception itself but rather in a particular view of man where human beings are seen as creatures whose anthropomorphic projections can be evoked "automatically" and their interaction with robots fully managed and controlled. From this perspective, the reason for using anthropomorphic design is to prove we are capable of mastering not only robots but also human beings. The alternative approach includes symbolic interactionism, i.e. one of the main perspectives in social science that stresses the human ability to actively construct meanings. The ultimate goal here is to shift from human-robot interaction to the human-robot-human paradigm.

Categories and Subject Descriptors
A. m [**General Literature**]: Miscellaneous

General Terms
Design, Theory

Keywords
Social Robots, Anthropomorphism, Meaning, Human-Robot-Human Paradigm

1. INTRODUCTION
In the pursuit of the perfect form for social robots we continue to ask about the degrees of anthropomorphism: To what extent should robots imitate human traits and which of them increase robot believability? What does it mean for a robot to be social? Does a robot need to be intelligent or it only needs to appear intelligent? These and other questions serve as frames of reference for robots that look and act like humans. The complexity and ever-changing nature of human beings, however, makes it very difficult to grasp the essence of humanness and express it via robot design. Yet, we eagerly face the challenge: As we explore new ways to answer the old question of what makes humans human, we believe we are closer than ever to finding the ultimate answer. We are quick to claim success in building robots that are our partners, social actors, emotional machines and so on. As we continue to anthropomorphise robots, we also risk denying people their human attributes (a phenomenon studied from different angles, which include mechanomorphism [3], technomorphism [7] and dehumanisation [8]). Obviously, this is not a goal to be purposefully pursued in robotics research. Objectification of man has been closely related to different processes taking place in our culture over the last three centuries, among which the emergence of modern science, capitalism and mass societies, as well as industrial revolution and the development of the technological mindset. As a result, the questions we ask now concern the degree of human-likeness in robots rather than whether it is possible to draw analogies between human beings and physical objects at all.

Social robots are not alive but they only *appear* as such. Thus, the anthropomorphic illusion relies not only on robot design but also on human perceptions and imagination. This is why we often attempt to control user experience the way it fits the anthropomorphic effect in robots. Such an approach is both a reason and a result of viewing anthropomorphism as a matter of automatic responses and people as organisms to be manipulated and played upon. This work argues instead that while our tendency to anthropomorphise is "a default schema" [3] intrinsic to human nature, there is always a degree of active interpretation involved in how we perceive and conceive robots. According to symbolic interactionism (one of the main perspectives in social science in general and in sociology in particular) [2], the distinctively human trait is the ability to construct meanings and act accordingly, where meanings are being constructed and negotiated in the course of social interaction. When creating the anthropomorphic illusion, social robotics too deals with the meanings people create and their subjective interpretation of robot appearance and behaviour. Thus, the potential of social robotics lies in bringing new insights into the human ability to make meanings through the new forms of social interaction with robots.

2. MANIPULATION
One could argue that building anthropomorphic robots always involves a degree of deception as social robots lead people into thinking machines are truly similar to humans (in other words, humanlike robots "trick" or "fool" people [4]). The negative versus positive effect of anthropomorphic deception is yet to be understood. The core of deception is manipulation, i.e. the act aimed at controlling and managing user experience, whether explicitly or not. In the line of this logic, we view the human tendency to anthropomorphise robots as "automatic" [5, 6] where robots are designed to give cues, elicit reactions and affect users through specific visual, audio and tactile stimuli.

Thus, anthropomorphic projections are often seen as a set of predefined reactions to be evoked by appropriate robot design. Such an approach allows translating user experience into terms and measures intelligible to robotic systems and defining universal principles for anthropomorphic robot design. It also constitutes an attempt to design and control not only robotics systems but also the entire process of human-robot interaction, users' performance included. At the same time, one of the most researched topics has been robot autonomy. The very definition of the social robot refers to "an autonomous or semi-autonomous robot" [1] and, despite the lack of the agreed framework for autonomy, building fully autonomous robots seems an inevitable step in robotics research. Thus, paradoxically, we place greater value on autonomy of robots than autonomy of humans. This is of crucial importance since how we conceptualise human beings has direct consequences on how we design social robots and what vision of man we promote through robot design.

Some researchers focus on purposefully deceptive robots. This is because deception is seen as a useful technique widely used by humans, and animals, that helps achieving specific goals. Also, many forms of manipulation are generally accepted in our society, in particular in the marketing industry. One could also argue that there is only a slight difference between deception and persuasion where "persuasive technologies" are meant to influence people's attitudes and behaviours, especially in the area of health and well-being. From this perspective, social robots are "persuasive agents" designed to influence people for their own good rather than merely deceive them. In any case, the main ethical issue here is a particular view of man where human beings are seen as creatures whose performance can be controlled, managed and optimised. It is worth asking about the underlying motivation: Why do we need to control and guide user experience? Many possible answers include that, while we play on fundamental human traits and needs, we aim to prove we are capable of mastering not only robots but also human beings.

3. INSPIRATION

As discussed above, one of the distinctively human traits is the ability to construct meanings, with social interaction being of crucial importance for the process of meaning-making (it is the nature of meanings that they are *shared* by individuals). Social robotics has human social interaction as the key frame of reference and one of the main goals for anthropomorphic robots. Yet, there is often little of the "social" or "shared" in human-robot interaction: Robots are typically designed to interact with single users or with only limited groups of people. On the one hand, this is largely due to technological constraints and the roles assigned to robots. On the other hand, such an approach is deeply rooted in contemporary Western culture, where the decline of social ties is becoming an increasing social problem (and the raison d'être for robot social companions). Social robotics risks following the approach where man is seen as an object to be isolated from any larger structure, controlled and evaluated according to the utility-based criteria ("useful" stands for "serviceable for an end on purpose" [Merriam-Webster], where the purpose refers to the successful achievement of the anthropomorphic illusion in robots). An alternative solution, however, includes inspiring new thinking in humans by means of anthropomorphic robots. The only way to advance our understanding of what actually makes us human is through the meanings that are constructed and modified by human users. This is why the goal for social robotics should lie in promoting the human-robot-human paradigm rather than only human-robot interaction. The understanding of human-robot-human interaction here is not limited to physical interaction or participation of more than one user. Instead, it implies broadening the social robotics perspective to see human-robot interaction as a catalyst for interactions between humans and further social inquiry.

4. CONCLUSIONS

If we agree that human nature is an inherently emergent process, then we shall agree that any attempt to petrify human traits in robots is doomed to fail. Yet, we are quick to claim success in creating robots similar to humans where we master not only robot but also human performance, be it through a genuine illusion or purposeful deception. This leads to objectification of man where the only way to make robots truly humanlike is to make humans less human. However, if we fully acknowledge the human ability to construct meanings exerted in the course of social interaction, social robots become powerful instruments to inspire new thinking on who we are as individuals and the human species. Therefore, rather than impose anthropomorphic projections on the robot users, the role for social robotics is to be a catalyst for human-human social interaction and human explorations on what it actually means to be human.

ACKNOWLEDGEMENTS

REFERENCES

[1] Bartneck, C. and J. Forlizzi. *A Design-Centred Framework for Social Human-Robot Interaction*. in *13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004)*.

[2] Blumer, H., *Symbolic Interactionism: Perspective and Method* 1986: University of California Press.

[3] Caporael, L.R., *Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine.* Computers in Human Behavior 1986. 2(3): p. 215-234.

[4] Duffy, B.R., *Anthropomorphism and The Social Robot.* Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems, 2003. 42(3-4): p. 177-190.

[5] Hegel, F., S. Gieselmann, A. Peters, P. Holthaus and B. Wrede. *Towards a Typology of Meaningful Signals and Cues in Social Robotics*. in *20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2011)*.

[6] Kiesler, S. *Fostering Common Ground in Human-Robot Interaction*. in *14th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2005)*.

[7] Vertesi, J., *"Seeing like a rover": Embodied Experience on the Mars Exploration Rover Mission*, in *CHI 2008 Extended Abstracts on Human Factors in Computing Systems* 2008, p. 2523-2532.

[8] Waytz, A., N. Epley and J.T. Cacioppo, *Social Cognition Unbound: Insights Into Anthropomorphism and Dehumanization.* Psychological Science 2010. 19(1): p.58-62.