



# The Self in the Machine

Karolina Zawieska<sup>\*,\*\*</sup>, Brian R. Duffy<sup>\*\*</sup>

\* Industrial Research Institute for Automation and Measurements PIAP, Warsaw

\*\* SMARTlab at University College Dublin (UCD), Dublin

**Abstract:** Social human-robot interaction (HRI) has been widely recognised as one of the major challenges for robotics. This social aspect includes designing robots that embraces various humanlike characteristics regarding robot appearance and behaviour. Few HRI studies, however, address the core goal of human social interaction, i.e. the development of the self. This paper argues that social robots can constitute an illusion or an extension of the human self and should be developed as such. The concept of the self and its meanings for HRI research are here discussed from the symbolic interactionism perspective.

**Keywords:** human-robot Interaction, self, self-extension, symbolic interactionism

**DOI:** 10.14313/PAR\_204/78

## 1. Introduction

As technology advances, social robots are embracing more and more humanlike qualities. Hence, the old question of what makes humans human gains a renewed focus. Amongst the various human characteristics that vary between different individuals and cultures, the self is a crucial trait inherent to all human beings that distinguish us from other living creatures. Thus, in order to fully embrace aspects of simulated human nature and reinforce more sophisticated human-machine interaction, social robotics need to address the concept of the self. So far, several attempts have been made to equip robots with a form of self-awareness, however, the social nature of the self and the way it arises out of embodied social interaction still needs to be addressed in greater detail.

## 2. The Human Self

Symbolic interactionism is one of the major sociological perspectives developed over the 20th century. It focuses on human social interaction seen as a process actively constructed and modified by individuals where the key element is the ability to give meanings to objects around them and communicate with others using symbols. Symbolic interactionist research focuses on a micro-level, however, inter-

action has been always analysed in a wider context: The main goal of human social interactions is to create and maintain the concept of the self as well as develop social structures based on a shared understanding of a given social and cultural reality. Depending on the approach developed within symbolic interactionism, the degree to which humans are influenced by or independent from social forces and other external and internal factors varies (for a detailed discussion on divergences between major theoretical and methodological assumptions see [36]). The central concept, however, remains unchanged: The notion of the self.

There is a large variety of theories analysing different components and functions of the self. The major assumption of symbolic interactionism is that the self “emerges in the social processes and is subject to change and modification throughout the individual’s life course” [18]. In other words, people develop self-conceptions and derive the feelings about themselves based on how others perceive them or how they imagine they appear to others. The key element in developing the self is the human reflexive ability, i.e. the ability to perceive the self as an object and interact with ourselves. Also, the development of the self requires the ability to use symbols and communicate with others where language is of crucial importance. Hence, the self and the social reality are being actively constructed and reconstructed within a given social and cultural framework. The majority of studies focus on how the self is constructed rather than what the content of the self is: “However defined, self refers to activity, to reflexive activity, and not to an object, thing, or essence” [33]. As society changes, the concept of self evolves too. The conceptions of self that were once anchored in stable spatial, institutional and temporal contexts are now anchored in impulses and multidimensional contexts, in particular those shaped by information technology. Due to the pace and multiplicitous modes of life, proliferation of communication technologies and growing instability and ephemerality of social structures necessary to anchor the self, identities and the self become fragmented, “saturated” and incoherent [12, 13, 31].

For the purposes of this work, it is important to emphasize that the self is what distinguishes the human from non-human agents and the ultimate goal of social interaction is to enable humans to fully develop the conception of

self. Although interactionists acknowledge the importance of biological predispositions that give people the potential to become human, they argue that people “develop into distinctively human beings as they take part in social interaction” [31]. Thus, social interaction is of crucial importance for human growth and survival not only as a biological entity but above all a human being. From this perspective, the study of social interaction between humans and robots addresses the essence of human nature.

### 3. The Robotic Self

From early stages, the notion of a robot was seen as a self-aware machine. In Capek’s “R.U.R.” play from which the term ‘robot’ comes, one of the robots named Radius is given a larger brain. As a result, the robot becomes self-aware and decides to refuse to work for humans [29]. As stated above, the sense of self is seen as a distinctive human trait, hence it is not surprising that the very first robots (as well as sculptures, dolls and anthropomorphic puppets that become “alive” in myths, literature and works of art) become self-aware. In Capek’s play, self-awareness is seen mostly as the result of the human brain activity and such an approach is present in many scientific studies [30]. Interactionists also recognise the importance of cognitive components (in addition to affective components [32]) in developing the concept of the self, however, they see both the conceptions of self and mind as social products [18, 36]. From the social robotics perspective, it is important to emphasize the embodied nature of the self. Symbolic interactionism typically focuses on language and communication as the main sources of the self. However, recent studies point at the awareness of the body and bodily sensations as an anchor for self-reflexivity [28]. There is no consensus on Cartesian dualism and many researchers see the self and the body as identical: “So, if the self is that which perceives, thinks, and acts, then the self is, above all, an embodied self” [4].

Social robotics constitutes a unique chance to study the concept of the self through embodiment and embodied social interaction. As embodied cognition aims to reinforce stronger integration between control system and the physical body of the robot, it inevitably develops the robots awareness of its body (even if only on an elemental level), hence constitutes the basis for development of the artificial self, reinforced by the development of social capabilities in machines. However, a variety of HRI studies focus on the degree and modes of anthropomorphism in robots, with only a few considering the self as a distinctive human trait and the ultimate goal of social interaction. Anthropomorphism is usually seen as the human tendency to perceive robots as humanlike in response to an (often limited) number of visual, audio or tactile stimuli provided by machines. The concept of anthropomorphism often remains weakly defined, or weakly managed. Obviously, only some human traits can be simulated in robots and it is a misconception to think that results from human-human interactions can be directly applied to human-robot interaction [6, 39]. Given the complexity and constructive nature of the

self, simulating the self in robots is a particularly difficult challenge that, however, should not be desisted.

### 4. Artificial Self-Awareness

Several researchers have included the self in their research in order to design robots that not only resemble human appearance and behaviour but are also defined as “self-aware” (self-reflexivity is the key element of the self). A common way to study such a phenomena is to create robots capable of visually recognising themselves in a mirror [25, 34] (such an idea has been inspired by the “mirror test” performed on chimpanzees [11]). On the other hand, it has also been argued that such a capacity is not enough for achieving self-awareness and the ability to attend to robot internal states is required [27]. Internal states or representations may include simulated emotions, beliefs and intentions and processes such as sensation, perception, conception, simulation, action, planning and thought. In such a case, a robot can be aware of its environments (external awareness) and of itself (internal awareness) [27]. Other researchers distinguish between internal and external attributes (language and processing capabilities, and physical features respectively). Awareness of such attributes contributes to robot identity seen as the ability to distinguish itself from other systems and environments [8]. The capacity to sense robot’s own internal states and react accordingly is also seen as indispensable for robots to “detect and deal with operational abnormalities”, i.e. to be self-aware [22]. Others define a self-aware system as a system that is “aware of itself thus able to monitor its own capabilities and limitations” [3]. In general, the common goal for creating self-aware robots is to enable them to better adapt to unpredictable environments and re-plan their behaviour [3, 27]. Whereas a few researchers have argued that “it is simplistic to define self-awareness as the ability to understand one’s own limitations, monitor one’s self, and detect and recover from faults” [23] and self-awareness needs to be “at least partially defined in terms of one’s awareness of others” [23]. One approach is to develop self-recognition in robots in order to endow robots with a capacity to learn about the concepts of “self” and “other” seen as useful in social interactions [15]. According to Hinds et al. [19] people may feel a sense of shared social identity with robots where the sense of shared identity is stronger with humanlike robots compared to machine-like robots. Some robots have been endowed with an artificial identity because of its importance for communication and emotional interaction with children [16], where identity is seen as a key element in order for the robot to become a believable character. A few researchers explicitly addressed the human ability to take the intentional stance towards themselves and actively construct the self. They argued that robots should “construct themselves as people” [23], i.e. develop identities and beliefs along with evolving and maintaining social relationships in order to become “persons”, where personhood does not require consciousness [23]. They also emphasized the role of the body in constructing human and robotic identities.

Since one of the primary goals for robotics is to create intelligent machines, artificial consciousness is often studied in relation to artificial intelligence (from the symbolic interactionist perspective too: “Rational conduct always involves a reflexive reference to self (...)” [24]). According to Minsky, humans possess only limited consciousness and its complexity should not be overestimated: “consciousness is actually easier to describe than most other aspects of mind” [26]. Also, Minsky views consciousness as preceding intelligence since “a certain degree of consciousness - in the sense of access to such records - is necessary for a person (or machine) to be intelligent. But even a large degree of such ‘consciousness’ would not by itself yield intelligence” [26]. On the contrary, others define the field of artificial consciousness as naturally emerging from AI research. Just like in AI and many other areas in robotics, it is possible to define consciousness in terms of weak and strong stance where the machine simulates or actually *is* conscious [5, 20]. For the purposes of this paper, it is important to emphasize that both artificial intelligence and artificial consciousness require embodiment and it is through the body and embodied social interactions that humans develop self-concepts. Intelligent robots that become or are self-aware, however, are sometimes feared as machines superior to humans in terms of intelligence but equal to them in terms of desire for independence or power that sooner or later lead machines to rebellion against their creators (for example, Radius from the “R.U.R.” play or HAL from 2001: A Space Odyssey – artificial intelligence embodied in the spacecraft). Thus, the goal to fully simulate or reproduce the human in the machine is severely challenged when it comes to creating autonomy in robots.

HRI researchers often demonstrate a certain degree of understanding of the importance of the self and identity for social interaction between humans and robots and the potential for the study of the self embodied social agents have. However, like many other concepts that follow the human frame of reference, defining the concept of the self remains challenging. In social sciences, there is no one standard of the self, however, it is clear that the self and identity are not the same: “In general terms, self is now viewed as a set or series of identities” [37], for example social, legal, bodily and digital identity. Also, in order to understand the self and identities, wider social contexts must be considered: “Because the self emerges in and is reflective of society, the sociological approach to understanding the self and its parts (identities) means that we must also understand the society in which the self is acting” [32]. Artificial self-awareness is therefore an inevitable evolution from work to date on developing socially capable embodied machines/robots to be used in real world settings.

## 5. Extension of the Self

It is interesting to note that several researchers emphasize the simulated nature of the self in robots, i.e. the fact that “a robot may not be truly self-aware even though it can have some characteristics of self-awareness” [27]. Some

researchers notice that it is difficult to predict whether machines could be conscious but “if a robot looks and appears to act human, it may be hard to resist treating it as a fully conscious person” [23]. This stands in contrast to approaches with many other human features, for example emotions or intelligence, that are often seen as possible to be not only imitated but also reproduced in robots. Such an approach towards self-awareness and consciousness constitutes a good basis for seeing humanlike robots as an extension of the human, in the sense where they simulate human traits (including the human self) rather than being artificial humans.

The idea that technology is an extension of the human being is not new. Social robots in particular reflect the anthropomorphic perspective that humans inevitably apply to all areas of their lives. As technology advances, the variety and complexity of human characteristics being simulated in robots increases. Telepresence robots are designed not only to simulate but also to extend the actions and sensations of the human operator in order to convey information to our sense of sight, hearing and touch [2]. From the symbolic interactionist perspective, material objects and symbolic meanings associated with them can serve as an extension of human identity, i.e. identity has boundaries that extend beyond the human body [7]. In particular, consumer identity is directly related to things, where “consumers use key possessions to extend, expand and strengthen their sense of self” [1]. By adopting a similar approach, people can extend not only their physical abilities but also their selves into machines, especially into humanlike robots that clearly evoke anthropomorphic projections. Some HRI researchers have also argued that people extend their self into robots and in general, into the objects they can control and personalise [17]. The control aspect is of crucial importance for the human perception of robots: as long as people exert control over machines they can extend their bodies and selves into robots, without the robot necessarily being anthropomorphic in appearance (for example, some Mars rover operators report they “look through their [the Rovers’] eyes” and “put [themselves] in the Rover’s head” [38]). As robot autonomy increases (or an illusion of autonomy), a distinction between an extension of the human and an independent agent becomes problematic.

Telepresence robots, especially humanoids, literally extend the human body. If we agree that the body and the self are indivisible, then we can assume that the extension of the body leads to the extension of the self. Thus, creating humanlike self and identity in the machine requires an active engagement with users and their imagination and the correct degree of anthropomorphism in robot design.

## 6. Illusion of the Self

The goal for social robotics is to create robots that can socially engage as much as possible. Given the fact that sociality is fundamentally a human trait, such a goal requires the use of the anthropomorphic form in robot design which may vary from abstract to realistic. Human

likeness can be achieved via purposefully anthropomorphic hardware and software design as well as via exploiting the human tendency to project human traits onto inanimate objects, either deliberately or not. Thus, social robots generate, to varying degrees, an illusion of human qualities rather than reproduce them, where the self and identity is no exception.

In order to embrace the relational nature of the human self, it is important to study it in the context of social interaction. Several studies addressed the topic of robots seen as relational artefacts and applied an interactionist perspective to social robotics [21, 35]. However, interactions between humans and robots are usually highly controlled, short-duration and limited to a single or small group of participants. As a result, robots are viewed as isolated objects and problem-solving is formal and narrowly defined leading to oversimplification of social interaction and related concepts (Forsythe [10] argued that such thinking is typical of AI researchers and it leads them to “delete the social”). Hence, the challenge lies not only in simulating the human being as a physical and psychological entity but also in incorporating robots into real and long-term social contexts characterised by large margins for interpretations and negotiation of meanings. Among all human traits simulated in robots, the self constitutes a key quality the illusion of which can help contribute to a perception of life in machines: “Maybe this is what we mean by artificial life, a *perceived* notion of consciousness artificially attained through anthropomorphism” [9].

## 7. Conclusions

The human self is fundamentally relational, i.e. developed in the course of social interaction. Social robotics and HRI research constitute a unique chance to better understand this aspect of the human nature: We can either simulate the self and identity in robots or let humans extend their own selves to the machine as well as redefine their self-concepts via the process of social interaction with anthropomorphic machines. In any case, the self in robots and their social characteristics are inevitably imitations or extensions of real human traits that make robots humanlike rather than human.

Understanding the concept of the self is particularly challenging in the context of current western culture, where identities are becoming more and more fragmented and blurred. Nevertheless, individuals are seen as fully responsible for their selves: “We are not what we are, but what we make of ourselves” [14]. Also, as virtual and augmented reality technologies continue to advance, the relationship between the body and identity is being continuously redefined. Thus, on one hand, social robots imitate human traits and challenge the confines of human nature. On the other hand, the embodied nature of robots and human-robot interaction can help verify our fuzzy conceptions of identity and anchor them in embodied social experiences. Perhaps paradoxically, it is the machine that can then help the human understand what the very core of being human actually is.

## Bibliography

1. Ahuvia A.C., *Beyond the Extended Self: Loved Objects and Consumers' Identity Narratives*. Journal of consumer research, 2005. **32**(1): 171–184.
2. Bar-Cohen Y., *Haptic Interfaces*, in *Automation, Miniature Robotics and Sensors for Nondestructive Evaluation and Testing*, A.S.f.N. Testing, Editor 2000.
3. Birlo M., Tapus A., *The Crucial Role of Robot Self-Awareness in HRI*, in *Proceedings of the 6th International Conference on Human-Robot Interaction* 2011, ACM: Lausanne, Switzerland. 115–116.
4. Cassam Q., *The Embodied Self*, in *The Oxford Handbook of The Self*, S. Gallagher, Editor 2011, Oxford University Press.
5. Chella A., Manzotti R., *Artificial Intelligence and Consciousness*. AI and Consciousness: Theoretical Foundations and Current Approaches, 2007.
6. Dautenhahn K., *Methodology & Themes of Human-Robot Interaction: A Growing Research Field*. International Journal of Advanced Robotic Systems, 2007. **4**(1): 103–108.
7. Dittmar H., *Material and Consumer Identities* [in:] *Handbook of Identity Theory and Research*, Schwartz S., Luyckx K., Vignoles V.L., Eds. 2011, Springer.
8. Duffy B.R., *Social Empowerment Of Autonomous Mobile Robotics*. in *1st International Conference on Autonomous Minirobots for Research and Edutainment* 2001. Paderborn, Germany.
9. Duffy, B.R., *Anthropomorphism and Robotics*, in *The Society for the Study of Artificial Intelligence and the Simulation of Behaviour* 2002: England.
10. Forsythe D.E., *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*. Writing Science 2002: Stanford University Press.
11. Gallup G.G., *Chimpanzees: Self-Recognition*. Science, 1970. **167**(3914): 86–87.
12. Gergen K.J., *The Saturated Self: Dilemmas of Identity in Contemporary Life* 1992: Basic books.
13. Gergen K.J., *Technology and the Self: From the Essential to the Sublime*. Constructing the Self in a Mediated Age, 1996: 127–140.
14. Giddens A., *Modernity and Self-identity: Self and Society in the Late Modern Age* 1991: Stanford University Press.
15. Gold, K. and B. Scassellati. *Learning About the Self and Others Through Contingency*. in *AAAI Spring Symposium on Developmental Robotics*. 2005.
16. Goris K., Saldien J., Vanderborght B., Lefeber D., *Probo, an intelligent huggable robot for HRI studies with children*, in *Human-Robot Interaction*, Chugo D., Editor 2010.
17. Groom V., L. Takayama P. Ochi and C. Nass, *I am my robot: the impact of robot-building and robot form on operators*, in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* 2009, ACM: California, USA. 31–36.
18. Herman N.J., Reynolds L.T., *Symbolic Interaction: An Introduction to Social Psychology* 1994: AltaMira Press.
19. Hinds, P.J., Roberts T.L., Jones H., *Whose job is it any-*

- way? *A study of human-robot interaction in a collaborative task*. Human-Computer Interaction, 2004. **19**(1): 151–181.
20. Holland, O., *Machine consciousness* 2003: Imprint Academic.
21. Jones R.A., *Relationalism through Social Robotics*. Journal for the Theory of Social Behaviour, 2013.
22. Lim W.Y. *Self-Awareness in Mobile Robots*. in *Robotics-DL tentative*. 1992. International Society for Optics and Photonics.
23. MacDorman K.F., Cowley S.J., *Long-term relationships as a benchmark for robot personhood*. in *The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006. IEEE.
24. Mead G., *Mind, Self, and Society: From the Standpoint of a Social Behaviorist (Works of George Herbert Mead, Vol. 1)* 1967: University Of Chicago Press.
25. Michel P., Gold K., Scassellati B., *Motion-Based Robotic Self-Recognition*. in *Proceedings. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004. IEEE.
26. Minsky M.L., *Conscious machines*. in *Machinery of Consciousness*. 1991.
27. Novianto R., Williams M.A., *The Role of Attention in Robot Self-Awareness*. *18th IEEE International Symposium on Robot and Human Interactive Communication*. 2009. IEEE.
28. Pagis, M., *Embodied Self-Reflexivity*. Social Psychology Quarterly, 2009. **72**(3): 265–283.
29. Richardson, L., *The Modern Robot and the Postmodern Cyborg: The Post-Human as an Image of Anxiety*. *Emergence: A Journal of Undergraduate Literary Criticism and Creative Research*, 2012. **3**.
30. Ross, J.A., *The Self: From Soul to Brain*. Journal of Consciousness Studies, 2003. **10**(2): 67-85.
31. Sandstrom, K., D. Martin and G. Fine, *Symbolic Interactionism at the End of the Century*, in *Hanbook of Social Theory*, G. Ritzer and B. Smart, Editors. 2001, SAGE Publications Ltd: London. 217–232.
32. Stets J.E., Burke P.J., *A Sociological Approach to Self and Identity*, in *Handbook of Self and Identity*, Leary M.R., Tangney J.P., Editors. 2005, The Guilford Press.
33. Stryker S., *Symbolic Interaction as an Approach to Family Research* [in:] *Symbolic Interaction: A Reader in Social Psychology*, Manis J.G. Meltzer B.N., Editors. 1978, AL-LYN AND BACON, INC.
34. Takeno, J., T. Takiguchi and A. Mizunaga, *A Study of Self-Awareness in Robots*. International Journal of Machine Consciousness, 2013. **5**(2): 145–164.
35. Turkle S., *A Nascent Robotics Culture: New Complicities for Companionship*, in *AAAI Technical Report Series* 2006.
36. Turner, J.H., *The Structure of Sociological Theory* 2003: Wadsworth Thomson Learning.
37. Turner, J.H., *Symbolic Interactionist Theories of Identity*, in *Contemporary Sociological Theory*, J.H. Turner, Editor 2012, SAGE Publications, Inc.
38. Vertesi, J., “*Seeing like a rover*”: embodied experience on the Mars exploration rover mission, in *CHI '08 Extended*

*ed Abstracts on Human Factors in Computing Systems* 2008, ACM: Florence, Italy. 2523–2532.

39. Zawieska K., Duffy B.R., Sprońska A., *Understanding anthropomorphisation in social robotics*. *Pomiary Automatyka Robotyka*, 2012(11): 78–82. ■

## Jażń maszyny

**Streszczenie:** Interakcja społeczna między ludźmi i robotami stanowi jedno z głównych wyzwań dla współczesnej robotyki. Aspekt społeczny interakcji między ludźmi i maszynami zawiera budowanie robotów, które posiadają cechy naśladujące cechy ludzkie w kontekście wyglądu i zachowania robota. Mimo to, niewiele prac skupia się na głównym celu ludzkich interakcji społecznych z robotami, tzn. na rozwinięciu jaźni w maszynie. Autorzy niniejszej pracy twierdzą, że roboty mogą stanowić rozszerzenie ludzkiej jaźni i jako takie powinny być rozwijane. Pojęcie jaźni oraz jego znaczenie dla badań w dziedzinie Human-Robot Interaction (HRI) omawiane są w pespektywy interakcjonizmu symbolicznego.

**Słowa kluczowe:** interakcja człowieka-robot, jaźń, rozszerzenie jaźni, interakcjonizm symboliczny

Artykuł recenzowany, nadesłany 06.11.2013 r., przyjęty do druku 21.01.2014 r.

### Karolina Zawieska, MSc

Karolina is a PhD student at SMARTlab in University College Dublin (UCD). She holds a Bachelor of Arts in Sociology from La Sapienza University of Rome and a Master of Arts in Applied Social Sciences from the University of Warsaw. Her PhD research explores anthropomorphism in machines: to what extent and why do people perceive robots as human-like and “alive”? She is a researcher at the Industrial Research Institute for Automation and Measurements PIAP in Warsaw, Poland.

e-mail: [kzawieska@piap.pl](mailto:kzawieska@piap.pl)



### Brian R. Duffy, PhD

Brian runs the Haptics & Robotics/HCI Research Group for SMARTlab in University College Dublin (UCD). He is also a senior engineer at SAP and has been actively involved in research in many international academic and non-academic institutions throughout Europe in the fields of robotics, artificial intelligence and haptics. Previously, he conducted postdoctoral research at University College Dublin, directed the Anthopos Group at Media Lab Europe, and undertook research for GMD, Germany and INSA de Lyon, France. He has a Masters of Engineering Science, a Bachelor of Science in Production Engineering, is a member of the IEEE, a Chartered Engineer, and holds the Eur. Ing qualification.

e-mail: [b.duffy@ucd.ie](mailto:b.duffy@ucd.ie)

