Abstract: This paper discusses different approaches to anthropomorphism in social robotics. While anthropomorphisation in robotics is often understood as the human tendency to perceive robots as humanlike in response to stimuli provided by machines, the authors argue that it is also an interpretative process which is highly subjective and limited in duration. To project human characteristics onto robots means to attribute qualities that robots do not have. Thus, in order to understand the nature of user engagement in human-robot interaction, it is of crucial importance to have a clear understanding of anthropomorphism in the field of social robotics.

Keywords: anthropomorphism, illusion of life, social robots, user engagement, Human-Robot Interaction

1. Introduction

Anthropomorphism, although basically understood in an abstract sense, is often perceived as a difficult topic to manage in the scientific and engineering world [41]. While its discussion in the field of social robotics varies greatly (as discussed later in this paper), a dictionary definition for anthropomorphism defines it as “an interpretation of what is not human or personal in terms of human or personal characteristics” [Merriam-Webster]. This paper focuses on anthropomorphism in social robotics, and inevitably on the animation of these artificial entities through a variety of technical means in order to (ideally) constructively exploit human-like characteristics in machines.

2. A matter of definition

Although widely observed, anthropomorphism is poorly understood and rarely investigated scientifically. The concept of anthropomorphism is closely related to the notion of the human nature (“When we understand the question ‘What is life?’ we will be able to build living machines” [9]), hence it is difficult to analyse it using scientific methods: “The vagueness of the concept means that one cannot indicate in any clear way the features of the thing to which the concept refers; hence, the testing of the concept by the empirical observation as well as revising the concept as a result of such observation are both made difficult. Because of this difficulty in effective validation such concepts are conducive to speculation in the unfavourable sense of that term…” [8]. In the field of social robotics researchers use the same terminology to describe humans and robots in terms of anthropomorphic characteristics (for example intelligent, autonomous or emotional) regardless of the fact that the ontological order of the two is radically different. It results in significant speculation with metaphors confused as statements and a lack of methodological validity.

Inevitably and involuntarily though, people will and do anthropomorphise [23]. Anthropomorphism has been the subject of much philosophical debate [2, 27, 37, 45] and it can be defined as the universal human tendency to ascribe human physical and mental characteristics to nonhuman entities, objects and events [10, 24]. It ranges “from spontaneous perception in daily life, to art, to science; from voices in the wind, to Mickey Mouse, to the Earth and Gaia” [23]. It is sometimes seen as “not the attribution of likeness, but its overestimation” [24]. Some of the qualities seem to be particularly important in creating the illusion of life, such as the human mind (closely related to agency) [10, 21, 22]. The type of characteristics projected onto objects and nonhuman agents depends on many factors, among which the individually and culturally shaped concept of humanness, age, gender and the context of use [16, 51].

According to the three-factor-theory of anthropomorphisation, variability in anthropomorphism can be explained through three determinants: elicited agent knowledge (use of existing knowledge about the self to make an inference about a nonhuman agent), social disconnection (search for connections with nonhuman agents due to a lack of social connection with other humans) and “effectance” motivation (use of familiar concepts to understand environments, including nonhuman agents) [16–18, 50]. Others define anthropomorphism in terms of its function, that is they argue that it can be seen either as a human cognitive default strategy (a peculiarity of human thinking) or a product of overlapping interspecies coordination systems (that enable the members of a given species to coordinate their own behaviour with that of their conspecifics) or a species-specific group-level coordination system (humans are obligatorily interdependent – social – hence they require a system for the coordination of behaviour where behaviour is being coordinated by values) [10].

Most human-robot interaction (HRI) researchers investigate either the extent to which people anthropomorphise robots based on specific human characteristics, or the accuracy in describing the actual capacities of non-human agents. Thus, anthropomorphisation in social robotics is usually understood as the human tendency to perceive robots as humanlike in response to visual, audio and/or tactile stimuli provided by robots. The main focus goes directly to measuring anthropomorphism while the concept of anthropomorphism itself remains largely undefined and unexplained. Bartneck et al. [5] reports on several measurements used for anthropomorphism among which a division of the concept into smaller, presumably better-known components such us sociability, intellect and personality measured with the help of a questionnaire [34],
a behavioural measurement that consists of analyzing differences in where the participants were looking when they looked at either a human or an android [38], and eventually “Godspeed questionnaires” developed by Bartneck et al. [5] that, among other items, measure anthropomorphism on a 5-point scale using the following semantic differential scales: Fake/Natural, Machinelike/Humanlike, Unconscious/Conscious, Artificial/Lifelike. Moving rigidly/Moving elegantly. Bartneck et al. are right to develop standardised tools for human-robot interaction and see the link between anthropomorphism and animacy, however, they do not define the terms they use to measure anthropomorphism. In most HRI studies to date, human-robot interaction is acultural and decontextualised (with few exceptions [4, 28, 44]) where “humans” constitute one category opposed and/or compared to robots – it leads to strong oversimplifications and does not increase the understanding of anthropomorphism.

In general, when studying anthropomorphism in social robotics, it depends on the way we conceptualise robots: if we define a robot as an artificial human being (the strong ontological claim according to which the humanoid actually becomes human [29]) anthropomorphic projections are taken literally; if we define a robot as a machine that only simulates a human being (the weak ontological claim where humanoid only appears to become human [29]) then we see anthropomorphism as a metaphor.

The concepts of weak and strong life in robots are related to our definition of a human being. All attempts to create robots that become human, have a tacit or explicit assumption that a human being can be reduced to the human body (psychological and internal experiences are viewed through their physical expressions) endowed with the qualities that can be measured, quantified and copied – not all of them, but enough to build an artificial human (surprisingly little has changed since the times when La Mettrie described a human being as a machine [12]). On the other hand, those who aim to build robots that only appear to be human, recognise the complexity of the human reality (made of the human body and the social and cultural context) where only some aspects can be only simulated.

Regardless of the field, anthropomorphism is still defined as the human tendency to project human qualities onto nonhuman entities. However, researchers differ on whether its function is positive or negative: according to one view, anthropomorphism distorts our understanding of animals and humans, i.e. it is a categorical mistake that should be avoided, in particular in ethology [10, 24, 26, 32], while others argue that a rejection of the natural human tendency to anthropomorphise leads to distortion [10]. Anthropomorphism can be also seen as a perceptual error, which is relatively easy if we mistakenly identify something as having human qualities, but significantly rewarding if an object of projected human characteristics proves to have them [24]. It is also possible that anthropomorphic thinking is “built into us” hence whether we should reject it or not is not an issue because “we could not abandon it even if we wished to” [32].

In social robotics, it is seen as a useful mechanism rather than a hindrance – it serves as a tool to create the interface between man and technology [13, 14], i.e. to bridge the gap between the living and inanimate (weak ontological stance) or to create an exact copy of a human being (strong ontological stance).

3. The concept of anthropomorphism

Given the variety of human characteristics, differences in individual imaginative predispositions, as well as the number of social contexts in which robots may be put to work, attempts to provide an exhaustive list of all humanlike qualities and degrees to which people may project them onto robots are doomed to fail. It is possible, however, to define the conceptual framework of anthropomorphisation:

a) Illusion

To project human characteristics onto robots means to attribute qualities that robots do not have (to make attributions which are not only unproven but also unlikely, i.e. to misattribute human traits [24, 26]). Simulating humans – by means of robot design or human interactions with the robot – is not equal to being human (the term anthropomorphism: “anthropos” – from Greek: man, human being, and “morphe” – from Greek: shape, form, refers to “human likeness” rather than to “humanness”). A robot is a machine – “a constructed thing whether material or immaterial” [Merriam-Webster]. It is an object and not an entity that uses the human tendency to anthropomorphise to give the illusion of life rather than becoming alive.

This is the reason why many researchers see anthropomorphism as a useful mechanism that facilitates human-robot interaction (in particular communication) and encourages familiarity with robots. Nonetheless, numerous researchers suggest that human interactions with robots endowed with humanlike abilities can be equal to human-human interactions: “Although clearly people attribute social properties to robots in research studies, it is less clear whether they believe that the robot literally possesses these characteristics... or whether they instead are using human terms metaphorically” [19] and also “Very little about these relationships [with robots] seemed to be experienced (as if)”. [49] Some researchers describe robots as perceived as “alive in some respects but not alive in other respects” [43] or “neither alive nor not alive” [30]. Others clearly state that social robots are “designed to create an illusion of real social interaction” but they also suggest “the use of anthropomorphic features can be a very effective way to accomplish meaningful social interaction between humans and robots” [31].

In social sciences some researchers also challenge the traditional understanding of social interaction being restricted to an exchange between humans and argue that new technologies have empowered computers, robots and avatars (and even “entities that reside in memories, projections, and imaginings – humans and nonhumans”) with critical interactive and communicative capacities that encourage humans “to perceive and react to these entities as legitimate partners in social interaction” (for a detailed discussion see [11]).

According to the authors of this paper, every element of social robotic systems, be it robot appearance, behaviour or the social context, generates, to varying degrees, an illusion of human physical, psychological and/or social qualities. While the nature itself of a user’s engagement with robots requires further research, it has however been recognised that “... the common everyday use, even among scientists, of anthropomorphic terms and metaphorical language can seem like animism if taken literally, and the failure to warn subjects to differentiate between scientific and metaphorical language devalues many studies” [25]. The fact that anthropomorphism is perceived as
too vague and unscientific highlights the difficulties in the field of social robotics where work inherently involves investigating the merging of human and machine in form and function.

b) Interpretation

According to the theory of symbolic interactionism, people do not automatically respond to external stimuli but act toward things on the basis of the meanings the things have for them (the meaning of things is formed in the context of social interaction with other people): “He [the human being] has to construct and guide his action instead of merely releasing it in response to factors playing on him or operating through him. He may do a miserable job in constructing his action, but he has to construct it” [8]. Anthropomorphism also involves an interpretative process: Robot design plays an important role in evoking human projections; however, it is the human being who actively creates and modifies meanings of robot appearance and behaviours. This simple view is often overlooked in HRI studies where the main focus goes to people’s responses to stimuli provided by robots, i.e. people’s perceptions of robots are seen as predefined reactions to be evoked by appropriate robot design.

Some researchers explicitly reject the targeting of complex human interpretative capacities: “Interactivity in the form of speech or gesture, especially, will prompt observers to anthropomorphise automatically, without any intent or thoughtful processing” [35]. The motivation underlying such research is primarily to fully control the user experience: Kiesler cites several studies [3, 40, 42] to support the idea that “people apply stereotypes and social heuristics, and enact social habits with interactive systems automatically and mindlessly” [35]. Kiesler also reports that the subjects in the experiment responded “not just mindlessly to the system’s appearance and behaviour, but also to their mental model of what the system represented” but she also suggests that “The solution is... to create in people’s minds an appropriate mental model of the robot automatically” [35].

Thus, the illusion given by the machine is seen as more powerful than the human ability to construct meanings of things they deal with. In the 18th century, electricity and magnetism were “prominently figured as a technical form of magic in the scientific imagination” [46]. Humanoid robotics based on the assumption that users automatically accept what they see as real can similarly be seen a new form of technical magic rather than science, primarily because of the difficulty evident in work to date in managing its non-scientific nature.

While the exploitation of human-like traits in social robotics, and specifically humanoids, inherently promotes a degree of immediate and experiential-based social interpretation, these are not new concepts and have been studied extensively — just not in the field of robotics [7, 47, 48]. Anthropomorphism has been widely used in puppets and dolls where the nature of people’s engagement is clearly defined as “the willing suspension of disbelief” or “make-believe games”: inanimate objects become alive because people choose to imagine them as such and make-believe it for a limited time, i.e. they actively contribute to creating the illusion of life while using anthropomorphic props (for a detailed discussion see [15]).

c) Temporality

One of the goals identified in social robotics is to build robots capable of establishing long-term relationships with humans [6, 20, 33] where “robots are no longer merely machines for achieving tasks but become social actors in real-world human environments” [1]. However, there has been no evidence so far of successfully maintaining such an illusion of life in robots over time. Anthropomorphism can sometimes last in beliefs about a diffuse causal agency in the universe (many religious believers endorse an anthropomorphic concept of God [10, 39]) but not in objects. This is because anthropomorphism is closely related to animism. To anthropomorphise means to project human-like life into nonhuman entities and objects. Life is the key to successful anthropomorphisation. There can be animism without anthropomorphism but there is no anthropomorphism without animism. No matter how anthropomorphic in appearance and behaviour, objects are not alive and the only way to give life to them is to create an illusion of life. Even though there is no scientific consensus on what a human being and life is, “[w]e all have an intuitive understanding of when something is alive or not” [9]. Hence, it is difficult to maintain the illusion of life in robots indefinitely.

The temporary nature of human interactions with robots is sometimes explained through “the novelty effect”, i.e. increased interest in the new technology that, however, fades over time. Some researchers suggest that “The main reasons behind this change are still unknown” [36] — considering the fact that it is usually the robot human-like appearance and behaviour that appeal to the human curiosity, one can explain the novelty effect through time constraints present in anthropomorphisation. Key elements in such cases of the novelty effect are:

- The initial exposure to the robot: First time
- Surprise in its abilities and human-like features
- Interpretation (whether correct or not) of its behaviours

Subsequently, there are clear reasons for the failure of the social robot in being able to sustain an illusion of life or suspend disbelief over time [15]. Unmanaged expectations are the key flaw in the design process. The initial expectation is often considerably greater than the robots actual capabilities, and is primarily due to its appearance and human-like characteristics.

4. Conclusions

As robots enter the human social space, HRI researchers need to deal with concepts and phenomena related to the social domain that are difficult to translate into engineering and computer science terms. They often lack conceptual and methodological consensus in the social sciences themselves. Anthropomorphisation is a perfect example. Our tendency to anthropomorphise machines should not be seen as a hindrance but rather, it being a powerful mechanism that can facilitate human-robot interaction, seems to be generally accepted. In order to be successfully implemented, however, the anthropomorphic paradigm should be used judiciously, clearly defined and controlled. Anthropomorphism is closely related to what it is to be a human being — the way people attribute human-like characteristics to robots reflects their vision of the human being (what makes a human being human). It is important to investigate those human qualities people project onto robots, as well as technical, individual and contextual factors that might evoke such projections. However, due to the complexity of human nature and the highly subjective nature of anthropomorphisation, such investigations will never be fully complete. In other words, it is impossible to provide a standard definition of humaneness; it is possible, however, to define a robot. As
long as we establish clear goals for social robotics and define robots in terms of what they are, what they can do, and not what they could be, the nature of the user engagement in human-robot interaction will also become clear. The end user should understand the robot as much as the designer should control that understanding.

Bibliography

Zrozumieć antropomorfizację robotów społecznych

Streszczenie: Artykuł przedstawia różne ujęcia antropomorfizmu w robotyce społecznej. Podczas gdy antropomorfizacja jest często rozumiana jako tendencja człowieka do przypisywania robotom cech ludzkich w odpowiedzi na bodźce dostarczone przez maszyny, autorzy niniejszej pracy twierdzą, że jest to także wysoce subiektywny proces interpretacyjny, którego trwanie jest ograniczone w czasie. Przypisywanie robotom właściwości ludzkich oznacza przypisywanie cech, których roboty nie posiadają. Z tego względu, dokładne zrozumienie antropomorfizacji w dziedzinie robotyki społecznej jest kluczowe dla zrozumienia sposobu, w jaki użytkownik angażuje się w interakcję z robotem.

Stowa kluczowe: antropomorfizacja, iluzja życia, roboty społeczne, zaangażowanie użytkownika, interakcja człowiek-robot

Katarzyna Zawieska, MSc
She is a PhD student at SMARTlab/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 Warsaw University. Her PhD research explores the illusion of life in machines: to what extent and why do people perceive robots as „alive”? She has been working at the Industrial Research Institute for Automation and Measurements PIAP since 2010. She is also involved in international research activities in the area of space robotics “MARS2013: Morocco Mars Analog Space Simulation”.

Brian R. Duffy, PhD
He 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 for over fifteen years. Previously, he conducted postdoctoral research at University College Dublin, directed the Anthropos Group at Media Lab Europe, and 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.

Agnieszka Sprońska, MSc
She holds a master degree in Polish Philology from Warsaw University and post-graduate diplomas in European studies and in management of promotion of science. Working in the Industrial Research Institute for Automation and Measurements PIAP since 2007, research specialist in the field of international projects, especially in the area of security and defence. Between 2008–2012 member of the TALOS Project coordination team, responsible for the management and dissemination activities, from 2010 parallelly involved in preparation and execution of the projects focused on the psychological and social factors of the human-machine cooperation. Currently, she is engaged in two FP7 Security projects as the workpackage leader, and in national project focused on the autonomy of mobile robots, as the co-manager of the consortium.

e-mail: aspronska@piap.pl

43. Severson R.L., Carlson S.M., Behaving as or behaving as if? Children’s conceptions of personified robots and the emergence of a new ontological category, “Neural Networks”, 2010, 23(8–9), 1099–1103.
47. Tillis S., Toward an Aesthetics of the Puppet: Puppetry As a Theatrical Art, 1992, Greenwood Press.