Anthropomorphism and the social robot

ARTICLE *in* ROBOTICS AND AUTONOMOUS SYSTEMS · MARCH 2003

Impact Factor: 1.26 · DOI: 10.1016/S0921-8890(02)00374-3 · Source: DBLP

1 AUTHOR:

Brian R. Duffy
Massachusetts Institute of Technology
61 PUBLICATIONS **606** CITATIONS

Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Robotics and Autonomous Systems 42 (2003) 177–190

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Anthropomorphism and the social robot

Brian R. Duffy*

Media Lab Europe, Sugar House Lane, Bellevue, Dublin 8, Ireland

Abstract

This paper discusses the issues pertinent to the development of a meaningful social interaction between robots and people through employing degrees of anthropomorphism in a robot's physical design and behaviour. As robots enter our social space, we will inherently project/impose our interpretation on their actions similar to the techniques we employ in rationalising, for example, a pet's behaviour. This propensity to anthropomorphise is not seen as a hindrance to social robot development, but rather a useful mechanism that requires judicious examination and employment in social robot research.

Keywords: Anthropomorphism; Social robots; Humanoid; Artificial intelligence; Artificial emotion

1. Introduction

With technology starting to provide us with quite robust solutions to technical problems that have constrained robot development over the years, we are becoming closer to the integration of robots into our physical and social environment. The humanoid form has traditionally been seen as the obvious strategy for integrating robots successful into these environments with many consequently arguing that the ultimate quest of many roboticists today is to build a fully anthropomorphic synthetic human. It should be noted that anthropomorphism in this paper is the rationalisation of animal or system behaviour through superposing aspects of the human observer and is discussed in depth in Section 3. The term robot refers to the physical manifestation of a system in our physical and social space, and as such, virtual characters and/or avatar-based interfaces are not discussed in this work.

This paper discusses the use of anthropomorphic paradigms to augment the functionality and behavioural characteristics of a robot (both anticipatory and actual) in order that we can relate to and rationalise its actions with *greater ease*. The use of human-like features for social interaction with people (i.e. [1–3]) can facilitate our social understanding. It is the explicit designing of anthropomorphic features, such as a head with eyes and a mouth that may facilitate social interaction. This highlights the issue that social interaction is fundamentally observer-dependent, and exploring the mechanisms underlying anthropomorphism provides the key to the social features required for a machine to be socially engaging.

1.1. The robot

The Webster's Dictionary defines a robot as: "any manlike mechanical being, by any mechanical device operated automatically, esp. by remote control, to perform in a seemingly human way." Through popular interpretations, this definition already draws associations between a robot and man. And in looking to define a social robot, the following has been proposed: "A physical entity embodied in a complex, dynamic, and social environment sufficiently empowered to

* Tel.: +353-1-474-2823; fax: +353-1-474-2809.
E-mail address: brd@media.mit.edu (B.R. Duffy).

0921-8890/03/$ – see front matter
doi:10.1016/S0921-8890(02)00374-3

behave in a manner conducive to its own goals and those of its community" [4].

There is a two-way interaction between a robot and that which it becomes socially engaged with, whether another robot or a person. It is in embracing the physical and social embodiment issues that a system can be termed a "social robot" [4]. This paper deals with human–robot interaction, which is proposed as the primary motivation for employing anthropomorphism in robotic systems (see [4] for explicit robot–robot social interaction).

A robot's capacity to be able to engage in meaningful social interaction with people inherently requires the employment of a degree of anthropomorphic, or human-like, qualities whether in form or behaviour or both. But it is not a simple problem. As discussed by Foner [5], people's expectations based on strong anthropomorphic paradigms in HCI overly increase a user's expectations of the system's performance. Similarly in social robotics, the ideal paradigm should *not* necessarily be a synthetic human. Successful design in both software and robots in HCI needs to involve a balance of illusion that leads the user to believe in the sophistication of the system in areas where the user will not encounter its failings, which the user is told not to expect. Making social robots too human-like may also defeat the purpose of robots in society (to aid humans). For example, perhaps a robot who comes across as too intelligent may be perceived as more selfish or as prone to weaknesses as humans, and thus not as desirable as a reliable entity. The issue of the social acceptance of robots is not discussed here, but is important in the greater debate regarding the context and degree of social integration of robots into people's social and physical space.

The social robot can be perceived as the interface between man and technology. It is the use of socially acceptable functionality in a robotic system that helps break down the barrier between the digital information space and people. It may herald the first stages where people stop perceiving machines as simply tools.

2. The big AI cheat

Are social robots, with embedded notions of identity and the ability to develop social models of those that it engages with, capable of achieving that age-old pursuit of artificial intelligence (AI)? Can the *illusion* of life and intelligence emerge through simply engaging people in social interaction? How much can this illusion emerge through people's tendency to project intelligence and anthropomorphise?

According to the MACHIAVELLIAN (or SOCIAL) INTELLIGENCE HYPOTHESIS, primate intelligence originally evolved to solve social problems and was only later extended to problems outside the social domain (recent discussion in [6]). Will AI effectively be achieved through robot "intelligence" evolving from solving social problems to later extending to problems outside the problem domain? If the robot "cheats" to appear intelligent, can this be maintained over time? Does it matter if it cheats? Is it important what computational strategies are employed to achieve this illusion? The principles employed in realising a successful "illusion of life" in Walt Disney's famous cartoon characters [7] have been well documented. While inspiration can be constructively drawn on how to apply similar strategies to designing social robots and create the illusion of life and intelligence, the problem for the functional design of the social robot is much more complex, of course, than cartoon characters as behind each character is a puppet master. While behind each robot lies a designer, the issue of physical embodiment and a robot's autonomy effectively distances the designer over time (analogous to the Classical AI failings demonstrated with the robot "Shakey" [8]). While the robot's form can be quite static after deployment, the design and, just as importantly, the maintenance of social behavioural functionality (inspired for example by Thomas and Johnston [7]) is by no means a trivial issue.

Proponents of **strong AI** believe that it is possible to duplicate human intelligence in artificial systems where the brain is seen as a kind of biological machine that can be explained and duplicated in an artificial form. This mechanistic view of the human mind argues that essentially understanding the computational processes that govern the brains characteristics and function would reveal how people think and effectively provide an understanding of how to realise an artificially created intelligent system with emotions and consciousness.

On the other hand, followers of **weak AI** believe that the contradictory term "artificial intelligence" implies that human intelligence can only be simulated.

An artificial system could only give the *illusion* of intelligence (i.e. the system exhibits those properties that are associated with being intelligent). In adopting this **weak AI** stance, artificial intelligence is an oxymoron. The only way an artificial system can become "intelligent" is if it cheats, as the primary reference is not artificial. A social robot could be the key to the great AI cheat as it involves another's perception and interpretation of its actions.

This paper follows the stance of weak AI where computational machines, sensory and actuator functionality are merely a concatenation of processes and can lead to an illusion of intelligence in a robot primarily through projective intelligence on the part of the human observer/participant (see a discussion of autopoiesis and allopoiesis regarding physical and social embodiment in [4]).

2.1. Projective intelligence

In adopting the weak AI stance, the issue will not be whether a system is fundamentally intelligent but rather if it displays those attributes that facilitate or promote people's interpretation of the system as being intelligent.

In seeking to propose a test to determine whether a machine could think, Alan Turing came up in 1950 with what has become well known as the Turing Test [9]. The test is based on whether a machine could trick a person into believing they were chatting with another person via computer or at least not be sure that it was "only" a machine. This approach is echoed in Minsky's original 1968 definition of AI as "[t]he science of making machines do things that would require intelligence if done by [people])".

In the same line of thought, Weizenbaum's 1960 conversational computer program Eliza [10], employed standard tricks and sets of scripts to cover up its lack of understanding to questions it was not pre-programmed for. It has proved successful to the degree that people have been known to form enough attachment to the program that they have shared personal experiences.

It can be argued that the Turing Test is a game based on the over-simplification of intrinsic intelligence and allows a system to have *tricks* to fool people into concluding that a machine is intelligent. But how can one objectify one's observations of a system and not succumb to such tricks and degrees of anthropomorphising and projective intelligence? Does it matter how a system achieves its "intelligence", i.e. what particular complex computational mechanisms are employed? Employing successful degrees of anthropomorphism with cognitive ability in social robotics will provide the mechanisms whereby the robot could successfully pass the age-old Turing Test for intelligence assessment (although, strictly speaking, the Turing Test requires a degree of disassociation from that which is being measured and consequently the physical manifestation of a robot would require a special variation of the Turing Test). The question then remains as to what aspects increase our perception of intelligence; i.e. what design parameters are important in creating a social robot that can pass this variation of the Turing test.

2.2. Perceiving intelligence

Experiments have highlighted the influence of appearance and voice/speech on people's judgements of another's intelligence. The more attractive a person, the more likely others would rate the person as more intelligent [11,12]. However, when given the chance to hear a person speak, people seem to rate intelligence of the person more on verbal cues than the person's attractiveness [12]. Exploring the impact of such hypotheses to HCI, Kiesler and Goetz [13] undertook experimentation with a number of robots to ascertain if participants interacting with robots drew similar assessments of "intelligence". The experiments were based on visual, audio and audiovisual interactions. Interestingly the results showed strong correlations with Alicke et al.'s and Borkenau's experiments with people–people judgements.

Such experimentation provides important clues on how we can successfully exploit elements of anthropomorphism in social robotics. It becomes all the more important therefore that our interpretations of what is anthropomorphism and how to successfully employ it are adequately studied.

3. Anthropomorphism

This paper has thus far loosely used the term anthropomorphism; however, it is used in different senses

throughout the natural sciences, psychology, and HCI. Anthropomorphism (from the Greek word *anthropos* for man, and *morphe*, form/structure), as used in this paper, is the tendency to attribute human characteristics to inanimate objects, animals and others with a view to helping us rationalise their actions. It is attributing cognitive or emotional states to something based on observation in order to rationalise an entity's behaviour in a given social environment. This rationalisation is reminiscent of what Dennett calls the *intentional stance*, which he explains as "the strategy of interpreting the behaviour of an entity (person, animal, artifact, whatever) by treating it *as if* it were a rational agent who governed its 'choice' of 'action' by a 'consideration' of its 'beliefs' and 'desires"' [14]. This is effectively the use of projective intelligence to rationalise a system's actions.

This phenomenon of ascribing human-like characteristics to non-human entities has been exploited in religion to recent animation films like "Chicken Run" (Aardman Animations, 2000) or "Antz" (DreamWorks & SKG/PDI, 1998).

Few psychological experiments have rigorously studied the mechanisms underlying anthropomorphism where only a few have seen it as worthy of study in its own right (for example [15–17]). The role of anthropomorphism in science has more commonly been considering it as a hindrance when confounded *with* scientific observation rather than an object to be studied more objectively.

Is it possible to remove anthropos from our science when we are the observers and build the devices to measure? Krementsov and Todes [18] comment that "the long history of anthropomorphic metaphors, however, may testify to their inevitability". If we are unable to decontextualise our perceptions of a given situation from ourselves, then condemning anthropomorphism will not help. Caporael [15] proposes that if we are therefore unable to remove anthropomorphism from science, we should at least "set traps for it" in order to be aware of its presence in scientific assessment.

Kennedy goes as far as to say that anthropomorphic interpretation "is a drag on the scientific study of the causal mechanisms" [19]. Building social robots forces a new perspective on this. When interaction with people are the motivation for social robot research, then people's perceptual biases have an influence on how the robot is realised. The question arising in this paper is not how to avoid anthropomorphism, but rather how to *embrace* it in the field of social robots.

Shneiderman [20] takes the extreme view of the role of anthropomorphism in HCI by stating that people employing anthropomorphism compromise in the design, leading to issues of unpredictability and vagueness. He emphasises the importance of clear, comprehensible and predictable interfaces that support direct manipulation. Shneiderman's comment touches on a problem which is not fundamentally a fault of anthropomorphic features, but a fault of the HCI designers in not trying to *understand* people's tendency to anthropomorphise, and thus they indiscriminately apply certain anthropomorphic qualities to their design which only lead to user over-expectation and disappointments when the system fails to perform to these expectations. The assumption has generally been that the creation of even crude computer "personalities" necessarily requires considerable computing power and realistic human-like representations. Since much of the resources tend to be invested in these representations, Shneiderman's sentiments seem more a response to the lack of attention to other equally important issues in the design. Such unmotivated anthropomorphic details may be unnecessary; as investigations show that using simple scripting of text demonstrates that "even minimal cues can mindlessly evoke a wide range of scripts, with strong attitudinal and behavioural consequences" [21].

Even if this were not the case, Shneiderman's argument is valid when the system in question is intended as a *tool* and not when the attempt is to develop a social intelligent entity. The question of how to develop AI in social robotics effectively requires human-like traits, as the goal is a human-centric machine. In the social setting, anthropomorphic behaviour may easily be as clear, comprehensible and predictable interfaces (and of course there must be a balance).

Moreover, Shneiderman's hesitancy with the role of anthropomorphism in HCI is based on obscuring the important distinction of metaphorical ascription of human-like qualities to non-human entities with the actual explanation of others' behaviours with human-oriented intentions and mental states. As Searle points out, there is an important, if subtle, difference in what he terms AS-IF INTENTIONALITY used

to rationalise, say, a robot's behaviour, and INTRINSIC INTENTIONALITY found in humans [22].

Nass and Moon [21] demonstrate through experimentation that individuals "mindlessly apply social rules and expectations to computers". Interestingly, the authors are against anthropomorphism as they base it on a belief that the computer is not a person and does not warrant human treatment or attribution. This highlights a predominant theme in HCI, which is to view anthropomorphism as "portraying inanimate computers as having human-like personality or identity" [23], or projecting intrinsic intentionality. This differs from the broader psychological perspective of anthropomorphism, which also includes metaphorically *ascribing* human-like qualities to a system based on one's interpretation of its actions. In this paper, anthropomorphism is a metaphor rather than an explanation of a system's behaviour.

The stigma of anthropomorphism in the natural sciences is similarly partly based on a rationalisation of animal or plant behaviour based on models of human intentionality and behaviour. The stigma does not stem from the appropriateness in describing the behaviour in terms of anthropomorphic paradigms but rather in cases when such paradigms are used as explanations of its behaviour. Such "explanations" are incorrect, but anthropomorphism is not restricted to only this. It also encompasses facilitation.

3.1. Anthropomorphism and robots

Consequently, social robots should exploit people's expectations of behaviours rather than necessarily trying to force people to believe that the robot has human reasoning capabilities. Anthropomorphism should not be seen as the "solution" to all human–machine interaction problems but rather it needs to be researched more to provide the "language" of interaction between man and machine. It can facilitate rather than constrain the interaction because it incorporates the underlying principles and expectations people use in social settings in order to fine-tune the social robot's interaction *with* humans.

The role of anthropomorphism in robotics in general should not be to build a synthetic human. Two motivations for employing anthropomorphism are firstly the design of a system that has to function in our physical and social space (i.e. using our tools, driving our cars, climbing stairs) and secondly, to take advantage of it as a mechanism through which social interaction with people can be facilitated. It constitutes the basic integration/employment of "humanness" in a system from its behaviours, to domains of expertise and competence, to its social environment in addition to its form. Once domestic robots progress from the washing machine and start moving around our physical and social spaces, their role and our dealings with them will change significantly. It is in embracing a balance of these anthropomorphic qualities for bootstrapping and their inherent advantage as machines, rather than seeing this as a disadvantage, that will lead to their success. The anthropomorphic design of human–machine interfaces has been inevitable.

It can be argued that the first motivation listed above does not constitute a strong argument as, for example, general wheelchair access could negate the necessity of highly energy consuming bipedal motion in robots. Similarly, due to the mechanistic capabilities of robots, the function or role of the robot in society should also be aimed at undertaking those actions that, as a machine, it is inherently good at. This paper embraces the second motivation of employing anthropomorphism, that of facilitating human–robot interaction.

As pointed out previously, to avoid the common HCI pitfall of missing the point of anthropomorphism, the important criterion is to seek a balance between people's expectations and the machines capabilities. Understanding the mechanisms underlying our tendencies to anthropomorphise would lead to sets of solutions in realising social robots, not just a single engineering solution as found in only designing synthetic humans (developed in Section 5).

Still the questions remain. Is there a notion of "optimal anthropomorphism"? What is the ideal set of human features that could supplement and augment a robot's social functionality? When does anthropomorphism go too far? Using real-world robots poses many interesting problems. Currently a robot's physical similarity to a full body person is only starting to embrace basic human-like attributes, and predominantly the physical aspect in humanoid research (see http://www.androidworld.com), but just as in HCI, rigorous research is needed to identify the physical attributes important in facilitating the social interaction. A robot not embracing the anthropomorphic paradigm in some form is likely to result in a persistent

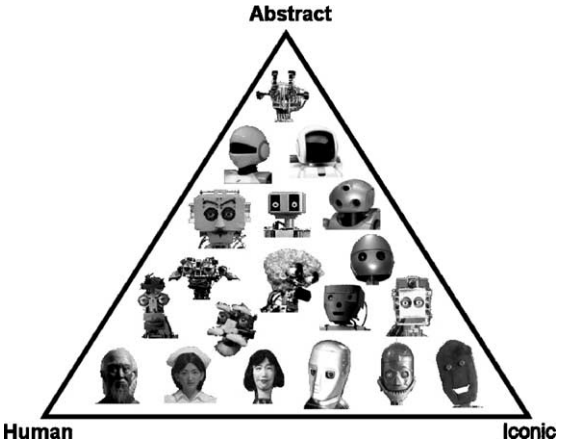

Fig. 1. Anthropomorphism design space for robot heads [26]. Notes: the diagram refers uniquely to the head construction and ignores body function and form. This is also by no means an exhaustive list. Examples were chosen to illustrate the proposed idea (motivated by McCloud [27]).

bias against people being able to accept the social robot into the social domain, which becomes apparent when they ascribe mental states to them. As shown in several psychological experiments and human–robot interaction experiments [13,24] and pointed out by Watt [25], familiarity may also ease social acceptance and even tend to increase people's tendency to anthropomorphise [16].

Fig. 1 provides an illustrative "map" of anthropomorphism as applied to robotic heads to date. The three extremities of the diagram (human, iconic and abstract) embrace the primary categorisations for robots employing anthropomorphism to some degree. "Human" correlates to an as-close-as-possible proximity in design to the human head. "Iconic" seeks to employ a very minimum set of features as often found in comics that still succeed in being expressive. The "Abstract" corner refers to more mechanistic functional design of the robot with minimal human-like aesthetics. An important question is how does one manage anthropomorphism in robots. One question is the degree and nature of visual anthropomorphic features. Roboticists have recently started to address what supplementary modalities to physical construction could be employed for the development of social relationships between a physical robot and people. Important arenas include expressive faces [1,2,28] often highlighting the importance of making eye contact and incorporating face and eye tracking systems. Examples demonstrate two methodologies that employ either a visually iconic [1,29] or a strongly realistic human-like construction (i.e. with synthetic skin and hair) [2] for facial gestures in order to portray artificial emotion states. The more iconic head defines the degree of anthropomorphism that is employed in the robot's construction and functional capabilities. This constrains and effectively manages the degree of anthropomorphism employed. Building mannequin-like robotic heads, where the objective is to hide the "robotic" element as much as possible and blur the issue as to whether one is talking to a machine or a person, results in effectively unconstrained anthropomorphism and a fragile manipulation of robot–human social interaction, and is reminiscent of Shneiderman's discontent with anthropomorphism.

Mori nicely illustrated the problematic issues found in developing anthropomorphic facial expressions on robotic heads [2] with "The Uncanny Valley" [30] (see Fig. 2). His thesis is that the more closely a robot resembles the human, the more affection it can engender through familiar human-like communication references. However, there is a region in the design space where the robot appears uncanny and weird. Issues of speed, resolution and expression clarification based on often very subtle actions provides for a highly complex design arena. The example facial expressions in [2] illustrate this problem.

Consequently, it can be argued that the most successful implementation of expressive facial features

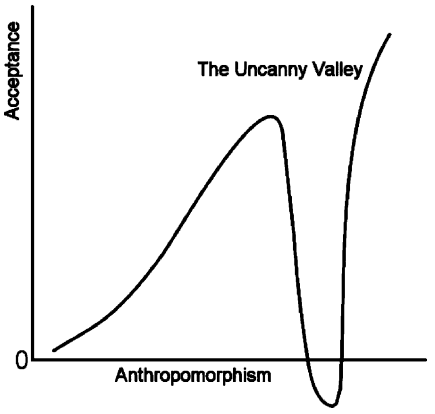

Fig. 2. Mori's "The Uncanny Valley".

is through more mechanistic and iconic heads such as Anthropos and Joe [29] and Kismet [1]. Strong human-like facial construction in robotics [2] has to contend with the minute subtleties in facial expression, a feat by no means trivial. Contrarily, it can be argued that successful highly human-like facial expression is an issue of resolution. Researchers are trying to bombard the artificially intelligent robot enigma with few looking for the key minimum features required to realise an "intelligent" robot. Consequently, in seeking to develop a social robot, the goal of many is the synthetic realistic human. From an engineering perspective it is more difficult to realise strong anthropomorphism as found in synthetic humans, i.e. [2], but in order to "solve" AI through hard research, it is a simpler more justifiable route to take. A more stringent research approach should not be to throw everything at the problem and force some answer, however, constrained, but rather to explore and understand the minimal engineering solution *needed* to achieve a socially capable artificial "being". The hard engineering approach is not mutually exclusive of the solution to realising the grail of AI. After all, engineering research plays a strong role in AI research (especially robotics), but the rest of AI research is why such engineering feats are needed. The path to the solution should embrace a more holistic approach combining both engineering solutions as well as tackling core AI research questions, even knowing what "cheats" to employ. A useful analogy is in not trying to replicate a bird in order to fly but rather recognising those qualities that lead to the invention of the plane.

From a social robotics perspective, it is not an issue whether people *believe* that robots are "thinking" but rather taking advantage of where people still have certain social expectations in social settings. If one bootstraps on these expectations, i.e. through exploiting anthropomorphism, one can be more successful in making the social robot less frustrating to deal with and be perceived as more helpful.

Anthropomorphising robots is not without its problems. Incorporating life-like attributes evokes expectations about the robot's behavioural and cognitive complexity that may not be maintainable [31].

Similarly it can be argued that this constrains the range of possible interpretations of the robot depending on the person's personality, preferences and the social/cultural context. This in fact can be reversed where constraining the robot to human-like characteristics can facilitate interpretation.

4. Artificial sociability

It is then natural to discuss sociality of robots. Artificial sociability in social robotics is the implementation of those techniques prevalent in human social scenarios to artificial entities (such as communication and emotion) in order to facilitate the robot's and people's ability to communicate.

The degree of social interaction is achieved through a developmental and adaptive process. The minimum requirement for social interaction is the ability in some way to adapt to social situations and communicate understanding of these situations through, for example a screen and keyboard or a synthetic speech system in order to have the "data-transfer" required. This functionality of communicating in addition to emotive expression, extends to the development and maintenance of complex social models and the ability to competently engage in complex social scenarios.

4.1. Communication and emotion

Communication occurs in multiple modalities, principally the auditory and visual fields. For artificial sociability the social robot does not necessarily need to be able to communicate information in as complex a manner as people but just sufficient enough for communication with people. For example, natural language is inherently rich and thus can be fuzzy, so strong anthropomorphism can have its limitations. However, the social robot needs to be able to communicate enough to produce and perceive expressiveness (as realised in speech, emotional expressions, and other gestures). This anthropomorphic ability can contribute to a person's increased perception of the social robot's social capabilities (and hence acceptance as a participant in the human social circle). Social competence is important in contributing to people's perception of another's intelligence [23]. This ability to understand the emotions of others is effectively the development of a degree of Emotional Intelligence in a robotic system. Daniel Goleman, a clinical psychologist, defines EI as "the ability to monitor one's own and others' emotions, to discriminate

among them, and to use the information to guide one's thinking and actions" [32].

Goleman discusses five basic emotional competencies: self-awareness, managing emotions, motivation, empathy and social skills in the context of emotional intelligence. This approach is a departure from the traditional attitude, still prevalent, that intelligence can be divided into the verbal and non-verbal (performance) types, which are, in fact, the abilities that the traditional IQ tests assess. While numerous often-contradictory theories of emotion exist, it is accepted that emotion involves a dynamic state that consists of cognitive, physical and social events.

As the case of Phineas Gage in 1848 demonstrated, an emotional capacity is fundamental towards the exhibition of social functionality in humans (see [33]). Salovey and Mayer [34] have reinforced the conclusions that emotionality is an inherent property of social intelligence in discussing emotional intelligence [33]. When coupled with Barnes and Thagard's [35] hypothesis that "emotions function to reduce and limit our reasoning, and thereby make reasoning possible", a foundation for the implementation of an emotional model in the development of a social robot becomes apparent. A social robot requires those mechanisms that facilitate its rationalisation of possibly overwhelming complex social and physical environments.

From an observer perspective, one could pose the question whether the attribution of artificial emotions to a robot is analogous to the Clever Hans Error [36], where the meaning and in fact the result is primarily dependent on the observer and not the initiator. Possibly not. As emotions are fundamentally social attributes, they constitute communicative acts from one person to another. The emotions have a contextual basis for interpretation, otherwise the communication act of expressing an emotion may not be successful and may be misinterpreted. The judicious use of the expressive social functionality of artificial emotions in social robotics is very important. As demonstrated through numerous face robot research projects (i.e. [2]), this treads a fine line between highly complex expressive nuances and anticipated behaviour.

4.2. Social mechanisms

What are the motivations that people could have in engaging in social interactions with robots? The need to have the robot undertake a task, or satisfy a person's needs/requirements promotes the notion of realising a mechanism for communication with the robot through some form of social interaction.

It has often been said that the ultimate goal of the human–computer interface should be for the interface to "disappear". Social robotics is an alternate approach to ubiquitous computing. The interaction should be rendered so transparent that people do not realise the presence of the interface. As the social robot could be viewed as the ultimate human–machine interface, it should similarly display a seamless coherent degree of social capabilities that are appropriate for interaction and to it being a machine. What matters is the ease and efficiency of communication with the robots, as well as the assistants' capabilities and persona.

This inherently necessitates that the robot be sufficiently empowered through mechanisms of personality and identity traits to facilitate a human's acceptance of the robots own mechanisms for communication and social interaction. It is important that the robot not be perceived as having to perform in a manner, which compromises or works against either its physical or social capabilities. This argues against the apparent necessity perceived by those building strong humanoid robots to develop a full motion, perception and aesthetic likeness to humans in a machine. Contrary to the popular belief that the human form is the ideal general-purpose functional basis for a robot, robots have the opportunity to become something different. Through the development of strong mechanistic and biological solutions to engineering and scientific problems including such measurement devices as Geiger counters, infra-red cameras, sonar, radar and bio-sensors, for example a robot's functionality in our physical and social space is clear. It can augment our social space rather than "take over the world". A robot's form should therefore not adhere to the constraining notion of strong humanoid functionality and aesthetics but rather employ only those characteristics that facilitate social interaction with people when required.

In order to facilitate the development of complex social scenarios, the use of basic social cues can bootstrap a person's ability to develop a social relationship with a robot. Stereotypical communication cues provide obvious mechanisms for communication between robots and people. Nodding or shaking of the head

are clear indications of acceptance or rejection (given a defined cultural context). Similarly, work on facial gestures with Kismet [1] and motion behaviours [3,37] demonstrate how mechanistic-looking robots can be engaging and can greatly facilitate ones willingness to develop a social relationship robots.

Another technique to create an illusion of "life" is to implement some form of unpredictability in the motion and behaviour of the robot to make it appear more "natural". Such techniques as Perlin Noise [38] have demonstrated the power of these strategies in animation, surface textures, and shapes.

Section 5 draws on the previous issues raised with a view to proposing a set of design criteria towards successfully employing anthropomorphism in social robots.

5. Achieving the balance

The previous sections have presented the issues pertinent to employing anthropomorphism in robotics, particularly for social interaction with people. The following criteria provide guidelines for those issues deemed key to the successful realisation of a robot capable of such social interaction. This list does not aim to describe the exact design methodologies for social robotics as the search for the ultimate social robot is not yet over, but rather provide useful concepts for anthropomorphism in social robot design methodology.

- *Use social communication conventions in function and form*. Social robotic research to date (between robots and people) has practically universally employed some human-like features (primarily the face, see Fig. 1). While the power of exploiting those conventions we are familiar with for communication is obvious, it is important to achieve the balance between mechanistic solutions and the human reference. An illustrative example would be a simple box on a table equipped with a speaker saying "no" and a robot head with two eyes turning to look straight at you and saying "no". The system should also be capable of reacting in a timely manner to its social stimulus and cues, whether external (from another) or internal (from desires, motivations, after a pause, etc.).
- *Avoid the "Uncanny Valley"*. Mori's "Uncanny Valley" [30] is becoming prevalent in building artificial systems aiming at conveying social cues and emotional interaction. The issue breaks down to an iconic vs. synthetic human problem. This relates the previous point. Up to a certain level, the more anthropomorphic features employed, the more the human participant has a sense of familiarity. Thus the robot's form facilitates social interaction. However, beyond this, combinations of particular anthropomorphic features can result in the reverse effect being induced. Mori maintains that the robot form should be visibly artificial, but interesting and appealing in appearance and effectively aim for the highpoint of the first peak in Fig. 2. A great many science fiction and fantasy manga and anime stories use this strategy.
- *Use natural motion*. Fluidity of motion has long been associated with the development of artificial personalities (i.e. Disney). Techniques such as Perlin Noise [38] can be implemented to combat the cliché.
- *Balance function and form*. There should be a strong correlation between the robot's capabilities and its form and vice versa. This helps combat ambiguity and misinterpretations about its abilities and role.
- *Man vs. machine*. Building a robot based on a human constrains its capabilities. The human form and function is not the ultimate design reference for a machine, *because it is a machine and not human*. This does not challenge our humanity, rather frees the robot to be useful to us. Trying to blur the boundary between a robot and human is unnecessary for a successful social robot. What is needed is a balance of anthropomorphic features to hint at certain capabilities that meet our expectations of a socially intelligent entity (and possibly even surprise and surpass these expectations). This is related to the "Uncanny Valley" as to what is enough, what is too much? What different sets of features will catalyse constructively our anthropomorphising tendencies? An important distinction should be drawn between *anthropomorphic features* and *anthropomorphic tendencies*. Features relate to the robot's form and function and tendencies relates to how these are perceived. Incorporating anthropomorphic features (human-like features) does not automatically facilitate familiarity or peoples tendencies to feel at ease with the robot's form and function. One needs to further explore the characteristics influencing this.

- *Facilitate the development of a robot's own identity*. Seamless integration into the social space and (nearly conversely) its differentiation through the robot's unique identity (which can be bootstrapped through stereotypical associations—see [4]) within that space to allow others build social models of the robot. Allowing the robot to portray a sense of identity makes it easier for people to treat it as a socially capable participant. In that sense, it is differentiated from the pool of social space. Because it is more a "participant", it is more seamlessly integrated into the social scene rather than simply existing physically somewhere in the social space.
- *Emotions*. Artificial emotional mechanisms [39–41] can guide the social interaction in (1) maintaining a history of past experiences (categorising experiences, i.e. memory organisation), (2) underlying the mechanism for instinctual reactions, (3) influencing decision making, especially when no known unique solution exists, (4) altering tones of communication as appropriate (as seen in characters in [42]), and (5) intensifying and solidifying its relationships with human participants. In order to convey some degree of emotional communication, familiar expressive features based on metaphors conventionally used in rationalising emotional expression can be employed. The judicious selection of artificial emotion generation techniques should facilitate the social interaction and not complicate the interaction. Here the use of a Facial Action Coding System (FACS) [43] based approach could provide a standardised protocol for emotional communication (see Kismet) for establishing stronger social relationships.
- *Autonomy*. The autonomy of the social robot is dependent on its social roles, capabilities and the requirements expected of it from both itself and others in its social environment. Autonomy also implies that the robot has to develop the capacity to interact independently and that its own capabilities and the social context in which it is situated allows it to do so. The issue of autonomy in robotics currently refers to battery life, the ability to self-navigate and other low-level interpretations.

In returning to Fig. 1, the most optimal system would be that which achieves a balance between human-like (familiarisation), iconic (heuristics), and functional (mechanistic solutions) features. This approximates roughly to the middle area of the anthropomorphism triangle.

Collectively, these mechanisms are based on a requirement to develop a degree of artificial sociability in robots and consequently artificial "life" and "intelligence". But if the robot becomes "alive", has science succeeded?

The following section briefly discusses how this notion of perceived robot "life" through artificial sociability will open a can of worms.

6. Fear of robots

"RADIUS: I do not want any master. I know everything for myself". Karel Capek's original robots from his 1920 play "Rossum's Universal Robots" were designed as artificial slaves which were intended to relieve humans of the drudgery of work. Unfortunately the story develops into the robots becoming more intelligent, becoming soldiers and ultimately wiping out the human race. Oops.

Asimov's laws from 1950 are inherently based on a robot's interaction with people. Interesting stories in his books "I, Robot" and "The Rest of the Robots" present scenarios where the often-bizarre behaviours of the robot characters are explained according to their adherence to these laws. An example is the telepathic robot that lies in order not to emotionally hurt people but inevitably fails to see the long-term damage of this strategy.

When the robot itself is perceived as making its own decisions, people's perspective of computers as simply tools will change. Consequently, issues of non-fear inducing function and form are paramount to the success of people's willingness to enter into a social interaction with a robot. The designer now has to strive towards facilitating this relationship, which may necessitate defining the scenario of the interaction. The physical construction of the robot plays an important role in such social contexts. The simple notion of a robot having two eyes in what could be characterised as a head, would intuitively facilitate where a person focuses when speaking to the robot. As highlighted in the PINO research, "the aesthetic element [plays] a pivotal role in establishing harmonious co-existence between the consumer and the product" [44]. Such physical attributes as size, weight,

proportions and motion capabilities are basic observable modalities that define this relationship (PINO is 75 cm tall). Children's toy design has for generations employed such non-fear inducing strategies for construction. An interesting example of fusing robotics and children's toys can be found in iRobot's "My Real Baby", where some reviews indicate that the fusion of children's expectations from a toy doll and the robotic aspect clash significantly. Roboticists have consequently to be careful to match behavioural expectations and the robot's form in social interaction with people.

Current research seeks to achieve a coherent synthesis between such often independent technologies as speech, vision, multi-agent systems, artificial emotions, learning mechanisms, localisation, building representations, and others. While the robot "Shakey" was the first well-known system to seek a cohesive system integration for the physical world, it highlighted some very important fundamental problems with Classical AI research, i.e. the physical embodiment issue [8]. Integrating robots into our social world (social embodiment [4]) will undoubtedly rear up some new dragons in AI and robotics research. While anthropomorphism in robotics raises issues where the taxonomic legitimacy of the classification 'human' is under threat, a number of viewpoints adopted by researchers highlight different possible futures for robotics: (a) machines will never approach human abilities; (b) robots will inevitably take over the world; (c) people will become robots in the form of cyborgs.

A fourth possibility exists. Robots will become a "race" unto themselves. In fact, they already have. Robots currently construct cars, clean swimming pools, mow the lawn, play football, act as pets, and the list is growing very quickly. Technology is now providing robust solutions to the mechanistic problems that have constrained robot development until recently, thereby allowing robots permeate all areas of society from work to leisure. Robots are not to be feared, but rather employed.

7. Future directions

Robots will expand from pure function as found in production assembly operations to more social environments such as responding and adapting to a person's mood. If a robot could perform a "techno handshake" and monitor their stress level through galvanic skin response and heart rate monitoring and use an infrared camera system measuring blood flow on the person's face, a traditional greeting takes on a new meaning. The ability of a robot to garner bio information and use sensor fusion in order to augment its diagnosis of the human's emotional state through for example, the "techno handshake" illustrates how a machine can have an alternate technique to obtain social information about people in its social space. This strategy effectively increases people's perception of the social robot's "emotional intelligence", thus facilitating its social integration.

The ability of a machine to know information about us using measurement techniques unavailable to us is an example where the social robot and robots in general can have their own function and form without challenging ours. What would be the role of the social robot? If therapeutic for say psychological analysis and treatment, would such a robot require full "humanness" to allow this intensity of social interaction develop?

The only apparent validation for building strong humanoid robots is in looking to directly replace a human interacting with another person. It is in seeking to embrace fundamental needs and expectations for internally motivated behaviour (for one's own means) that could encourage strong humanoid robotics. Examples include situations requiring strong introspection through psychotherapy or the satisfaction/gratification of personal needs or desires such as found in the adult industry.

Does work on the development of socially capable robots hark a fundamental change in AI and robotics research where work is not based on a desire to understand how human cognition works but rather on simply designing how a system may appear intelligent? The answer is obviously no. In order to design a system that we could classify as "intelligent" we need to know what intelligence is, and even better if we can understand what it is.

8. Conclusion

It is argued that robots will be more acceptable to humans if they are built in their own image. Arguments

against this viewpoint can be combated with issues of resolution where its expressive and behavioural granularity may not be fine enough. But if the robot was *perceived* as being human, then effectively it should be human and not a robot. From a technological standpoint, can building a mechanistic digital synthetic version of man be anything less than a cheat when man is not mechanistic, digital nor synthetic? Similar to the argument to not constrain virtual reality worlds to our physical world, are we trying to constrain a robot to become too animalistic (including humanistic) that we miss how a robot can constructively contribute to our way of life?

The perspectives discussed here have tended towards the pragmatic by viewing the robot as a machine employing such human-like qualities as personality, gestures, expressions and even emotions to facilitate its role in human society. The recent film AI (Warner Brothers, 2001) highlights the extended role that robot research is romantically perceived as aiming to achieve. Even the robot PINO [44] is portrayed as addressing "its genesis in purely human terms" and analogies with Pinocchio are drawn, most notably in the name. But, does the robot *itself* have a wish to become a human boy? Or do we have this wish for the robot? We should be asking why.

Experiments comparing a poor implementation of a human-like software character to a better implementation of a dog-like character [37] promotes the idea that a restrained degree of anthropomorphic form and function is the optimal solution for an entity that is not a human.

It can be argued that the ability of a robot to successfully fool a person into thinking it is intelligent effectively through social interaction will remain an elusive goal for many years to come. But similarly *machine* intelligence could already exist, in "infant" form. Robots are learning to walk and talk. The issue is rather *should* they walk? Walking is a very inefficient motion strategy for environments that are more and more wheel friendly. Should robots become the synthetic human? Is attacking the most sophisticated entity we know, ourselves, and trying to build an artificial human the ultimate challenge to roboticists? But researchers will not be happy building anything less than a fully functional synthetic human robot. It is just not in their nature.

There is the age-old paradox where technologists predict bleak futures for mankind because of their research directions but nevertheless hurtle full steam ahead in pursuing them. Strong humanoid research could be just a means to play God and create life. But fortunately, robots will not take over the world. They will be so unique that both our identity and theirs will be safe. As an analogy, nuclear energy has provided both good and bad results. We similarly need to not let one perspective cloud the other in social robotics and robotics as a whole.

While anthropomorphism is clearly a very complex notion, it intuitively provides us with very powerful physical and social features that will no doubt be implemented to a greater extent in social robotics research in the near future. The social robot is the next important stage in robot research and will fuel controversy over existing approaches to realising artificial intelligence and artificial consciousness.

References

[1] C. Breazeal, Sociable machines: expressive social exchange between humans and robots, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 2000.
[2] F. Hara, H. Kobayashi, Use of face robot for human–computer communication, in: Proceedings of the International Conference on Systems, Man and Cybernetics, 1995, p. 10.
[3] B. Duffy, G. Joue, J. Bourke, Issues in assessing performance of social robots, in: Proceedings of the Second WSEAS International Conference, RODLICS, Greece, 2002.
[4] B. Duffy, The social robot, Ph.D. Thesis, Department of Computer Science, University College Dublin, 2000.
[5] L. Foner, What's agency anyway? a sociological case study, in: Proceedings of the First International Conference on Autonomous Agents, 1997.
[6] H. Kummer, L. Daston, G. Gigerenzer, J. Silk, The social intelligence hypothesis, in: P. Weingart, P. Richerson, S. Mitchell, S. Maasen (Eds.), Human by Nature: Between Biology and Social Sciences, vol. 47, Lawrence Erlbaum Associates, Hillsdale, NJ, 1991, pp. 67–81.
[7] F. Thomas, O. Johnston, The Illusion of Life: Disney Animation, revised edition, Hyperion, 1995.
[8] N. Nilsson, Shakey the robot, Technical Note 323, SRI A.I. Center, April 1984.
[9] A. Turing, Computing machinery and intelligence, Mind 59 (1950) 433–460.
[10] J. Weizenbaum, ELIZA—a computer program for the study of natural language communication between man and machine, Communications of the ACM 9 (1966) 36–45.

[11] M. Alicke, R. Smith, M. Klotz, Judgements of physical attractiveness: the role of faces and bodies, Personality and Social Psychology Bulletin 12 (4) (1986) 381–389.
[12] P. Borkenau, How accurate are judgements of intelligence by strangers?, in: Proceedings of the Annual Meeting of the American Psychological Association, Toronto, Ont., Canada, 1993.
[13] S. Kiesler, J. Goetz, Mental models and cooperation with robotic assistants. http://www-2.cs.cmu.edu/~nursebot/web/papers/robot_chi_nonanon.pdf.
[14] D. Dennett, Kinds of Minds, Basic Books, New York, 1996.
[15] L. Caporael, Anthropomorphism and mechanomorphism: two faces of the human machine, Computers in Human Behavior 2 (3) (1986) 215–234.
[16] T. Eddy, J.G.G. Gallup, D. Povinelli, Attribution of cognitive states to animals: anthropomorphism in comparative perspective, Journal of Social Issues 49 (1993) 87–101.
[17] P. Tamir, A. Zohar, Anthropomorphism and teleology in reasoning about biological phenomena, Science Education 75 (1) (1991) 57–67.
[18] N. Krementsov, D. Todes, On metaphors, animals, and us, Journal of Social Issues 47 (3) (1991) 67–81.
[19] J. Kennedy, The New Anthropomorphism, Cambridge University Press, Cambridge, 1992.
[20] B. Shneiderman, A nonanthropomorphic style guide: overcoming the humpty–dumpty syndrome, The Computing Teacher 16 (7) (1989) 5.
[21] C. Nass, Y. Moon, Machines and mindlessness: social responses to computers, Journal of Social Issues 56 (1) (2000) 81–103.
[22] J. Searle, Intentionality: An Essay in the Philosophy of Mind, Cambridge University Press, New York, 1983.
[23] K. Isbister, Perceived intelligence and the design of computer characters, in: Proceedings of the Lifelike Characters Conference, Snowbird, UT, 1995.
[24] M. Gill, W. Swann, D. Silvera, On the genesis of confidence, Journal of Personality and Social Psychology 75 (1998) 1101–1114.
[25] S. Watt, A brief naive psychology manifesto, Informatica 19 (1995) 495–500.
[26] Robot heads (Fig. 1, from top to bottom-right): COG:MIT-AI Lab, http://www.ai.mit.edu; SIG: Kitano Symbiotic Systems Project, http://www.symbio.jst.go.jp/sigE.htm; ASIMO: Honda http://www.honda.co.jp/ASIMO/; The Humanoid Cranium Robot: Waseda University, http://www.humanoid.waseda.ac.jp; H6: JSK Laboratory, http://www.jsk.t.u-tokyo.ac.jp/research/h6/; SDR4X: Sony, http://www.sony.com.au/aibo/; Kismet: MIT AI Lab http://www.ai.mit.edu/projects/sociable/; JoeRobot: Media Lab Europe, http://anthropos.mle.ie; Isamu: Kawada Industries Inc., http://www.kawada.co.jp/ams/isamu/index_e.html; Inkha: King's College London, http://www.inkha.net; Doc Beardsley: CMU, http://www.etc.cmu.edu/projects/iai/; Elvis: Chalmers University of Technology, humanoid.fy.chalmers.se; Hadalay-2: Waseda University, http://www.humanoid.waseda.ac.jp/; Master Lee: YFX Studio, http://www.yfxstudio.com/human.htm; Saya: Kobayashi Lab, koba0005.me.kagu.sut.ac.jp/newsinfo.html; Roberta: Science University of Tokyo, hafu0103.me.kagu.sut.ac.jp/haralab/; R.Max: http://www.howtoandroid.com; Maxwell: Medonis Engineering, http://www.medonis.com; Woody: Media Lab Europe, anthropos.mle.ie.
[27] S. McCloud, Understanding Comics: The Invisible Art, Kitchen Sink Press, 1993.
[28] A. Cozzi, M. Flickner, J. Mao, S. Vaithyanathan, A comparison of classifiers for real-time eye detection, in: Proceedings of the ICANN, 2001, pp. 993–999.
[29] B. Duffy, The anthropos project. http://www.mle.ie/research/socialrobot/.
[30] M. Mori, The Buddha in the Robot, Charles E. Tuttle Co., 1982.
[31] D. Dryer, Getting personal with computers: how to design personalities for agents, Applied Artificial Intelligence 13 (3) (1999) 273–295. Special Issue on "Socially Intelligent Agents".
[32] D. Goleman, Emotional Intelligence, Bantam Books, 1997.
[33] A. Damasio, Descartes' Error: Emotion, Reason, and the Human Brain, G.P. Putnam's Sons, New York, 1994.
[34] P. Salovey, J. Mayer, Emotional intelligence, Imagination, Cognition, and Personality 9 (1990) 185–211.
[35] A. Barnes, P. Thagard, Emotional decisions, in: Proceedings of the 18th Annual Conference of the Cognitive Science Society, Erlbaum, 1996, pp. 426–429.
[36] O. Pfungst, Clever Hans (The Horse of Mr. von Osten): A Contribution to Experimental Animal and Human Psychology, Holt, Rinehart & Winston, New York, 1965.
[37] S. Kiesler, L. Sproull, K. Waters, A prisoner's dilemma experiment on cooperation with people and human-like computers, Journal of Personality and Social Psychology 70 (1) (1996) 47–65.
[38] K. Perlin, An image synthesizer, in: Computer Graphics, SIGGRAPH'85 Proceedings, vol. 19, 1985, pp. 287–296.
[39] R. Plutchik, A general psycho-evolutionary theory of emotion, in: R. Plutchik, H. Kellerman (Eds.), Emotion: Theory, Research, and Experience, vol. 1, Academic Press, New York, 1980, pp. 3–33.
[40] F. Michaud, E. Robichaud, J. Audet, Using motives and artificial emotions for prolonged activity of a group of autonomous robots, in: Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI 2001 Fall Symposium, AAAI Technical Report FS-01-02, MA, 2001.
[41] L. Botelho, H. Coelho, Machinery for artificial emotions, Cybernetics and Systems 32 (5) (2001) 465–506.
[42] K. Isbister, B. Hayes-Roth, Social implications of using synthetic characters: an examination of a role-specific intelligent agent, Tech. Rept. KSL-98-01, Knowledge Systems Laboratory, January 1998.
[43] P. Ekman, E. Rosenberg (Eds.), What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS), Oxford University Press, New York, 1997.
[44] F. Yamasaki, T. Miyashita, T. Matsui, H. Kitano, PINO the humanoid: a basic architecture, in: Proceedings of the Fourth International Workshop on RoboCup, Melbourne, Australia, 2000.

Brian Duffy is a Research Associate at the Media Lab Europe, the European research partner of the MIT Media Lab. He received his B.S. in Engineering from Trinity College, and his Masters of Engineering Science and Ph.D. in Computer Science from the University College Dublin. From 1992 to 1994, he worked at the National Institute for Applied Science in Lyon. From 1994 to 1996, he conducted research in artificial intelligence and built robot prototypes at the Fraunhofer-Gesellschaft's Institute for Autonomous Intelligent Systems. His current research aims to develop an amicable anthropomorphic socially capable office robot.