Anthropomorphism and Robotics

Brian R. Duffy

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

brd@media.mit.edu

Abstract

Should a robot be constrained to look like a human? A difficult question, but extremely relevant at a stage where technology is starting to provide us with some quite robust solutions to technical problems that have constrained robot development over the years. This paper discusses the role that anthropomorphism plays in robotics research and proposes that research on human-robot interaction or social robotics is the primary motivation for incorporating human aspects in robotics.

1 Introduction

Robotic research has evolved through a few basic stages when considering the impact of computers as control mechanisms. The most notable example of the application of Classical AI in the late 1960's was the robot Shakey (Nilsson, 1984). This highlighted the inability of computationally intensive centralised control systems to deal with real-world unpredictability and complexity. A rebelling against the strongly deliberative approach resulted in highly reactive techniques where little to no internal representation of the environment was maintained (Brooks, 1986). Issues regarding the difficulty in realising explicit goals have limited the development of purely reactive control approaches. In recent years, one could say that following the emergent intelligence experiments with multiple simple reactive robots (Cao *et al.*, *1997*), the idea of developing more explicit social functionality between a collective of robots gained momentum (Duffy, 2000). While humanoid robotics has arguably been *the* ultimate robot research, the implications of this social dimension to robotics have reinforced the design and building of human-like robots (Honda Asimo, PINO). In the extreme, the building of strongly human-like robots constrains the functionality that a robot could otherwise have. The role of anthropomorphism in robotics is not to build an artificial human, but rather to take advantage of it as a mechanism through which social interaction can be facilitated. Once robots come out of 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 their inherent advantage in being machines rather than seeing this as a disadvantage that will lead to their success.

2 The Humanoid

However disparate the means by which humanoid robot research has evolved, most are concerned with the human form and corresponding motor and perceptual functionality required to realise a physical and functional anthropomorphic robot. The exterior design of the robot has therefore aimed to define its functionality and facilitate its assertion of its identity as distinct from that of a mere object. Such research in building an artificial human has been fundamentally driven by two distinct motivations:

1. Addressing the engineering issues of building an artificial entity capable of performing in people environments with a similar capacity to a human
2. Building mechanisms whereby computational models can be implemented and tested in order to better understand human beings.

These areas stem from either the perspective of being inspired by humans in order to realise sophisticated machines, or use machines to understand humans. Research centring on the development of a humanoid machine from a mechanical engineering approach (Honda Asimo, PINO) have lead to impressive results. Similarly, the development of symbolic machines by which artificial mechanisms of "thought" are developed are still playing "catch-up" with the human brain. Yet a freely moving autonomous entity reacting to diverse sensory information through control mechanisms, with a sufficient degree of complexity, can still simulate a degree of intelligent behaviour.

It is proposed that success in designing a robot's functionality is not to attempt to make it as multi-functional and versatile as ourselves, but rather to employ those characteristics and attributes that are intrinsically "machine-like" to address capabilities that we do not have and require machines to do. It is important to make the distinction between the issues of whether one views the human body as inadequate as could be perceived in aspects of cyborg research, or if one aims to incorporate functionality on a robot that we simply require machines to perform anyway.

An interesting third viewpoint is to let the robot "be all it can" (i.e. employ technologies that provide functionality that humans do not have) with the humanlike features contributing to the social scenarios that arise when this robot is interacting with people, rather than being the paradigms through which the robot is constrained in its functionality and appearance. This social interaction paradigm is the only real justification for building robots that look like humans. The ascription of human features and metaphors, to facilitate interaction between man and machine in social scenarios, are important. Work to date on investigating the social interaction between people and robots employing human face-like features have demonstrated the importance of such analogies (Hara & Kobayashi, 1995; WE-3RIV, 2001; Brezeal, 2000). On the other hand, if functional requirements for a robot *as a machine* outside any social context, efficient design would most likely require the physical form of the robot to constitute something completely different from the human form.

The general objective of Social Robot research is to build a robot that can engage in social interaction scenarios, often with humans, in a familiar and compelling manner providing social communicative functionality that is natural and intuitive. From a scientific perspective, the use of such terms as "familiar", "compelling", "natural" and "intuitive" are about as difficult to deal with as the notion of "anthropomorphism" (see section 4). It is the psychological affiliation with the object that presents interesting challenges. This paper argues that these underlying perceptions hint at characteristics for a social robot, which eases the engineering task of creating flawlessly realistic humanoids. The *Placebo* project (Tony Dunne & Fiona Raby) demonstrates the development of personal attachments to objects and has effectively already highlighted the magnitude of flexibility available for robots strongly endowed with social functionality and features.

2.1 Human-robot relations

Recent robot research has studied the relations between humans and robots (Breazeal, 2000; Iida *et al.*, 1999; Duffy, 2002). These relationships will enable humans to interact meaningfully with a robot. The designer now has to strive towards facilitating the relationship between the human and the robot often with a view towards 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" (PINO). Such physical attributes as size, weight, proportions and motion capabilities are basic observable modalities that define this relationship (PINO is approximately 75cm tall).

Anthropomorphism is prevalent in robotics because of one's tendency to need such familiarity. Robotics research focuses on building humanoids not because a humanoid is the most efficient design for *any* given task, but rather because of one's tendency and need to anthropomorphise. Searle already points out the dangers present through confounding reasoning and rationalising because of this tendency to anthropomorphise (1992). Effectively, anthropomorphism obstructs the fact that the human form is not the ideal for a machine. However, objectively, the humanoid form best helps us interact with machines *because* of the fact that we anthropomorphise. But this does not necessarily mean that one should build seamlessly perfect-looking human-machines as anthropomorphism allows roboticists be heuristic in robot design (i.e. cartoon characters).

And then the robot looks *back* at you. This adds a further dimension to the interaction. Maybe this is what we mean by artificial life, a *perceived* notion of consciousness artificially attained through anthropomorphism. This intimate social scenario between a human and a physical robot will become a very important issue in social robot development.

3 The Role of the Social Robot

What role will such a robot have? While a dreaded question, the answer is actually quite straightforward. It is not the definition of a specific task that warrants the research, but rather any situation that requires people to interact with a machine. The application scenarios that naturally stem from this can include a museum guide, a corporate building butler, or a post office clerk. In fact, the fundamental issue is to ascertain what the robot, *as a machine*, is inherently *good* at.

Some argue that the design and construction of strong humanoid form and functionality is such that robots can be successfully integrated into human society where their primary mobile functions will be to negotiate the same obstacles encountered by humans in daily life. This is a weak argument when the proliferation of wheelchair access can minimise the necessity for two legs over the more efficient wheeled chassis for a robot. Even work in hazardous environments such as nuclear plants does not require a human friendly set-up but rather can have a "robot-friendly" configuration for highly specialised machines.

Contemporary interaction between humans and technology is dictated by machine-specific operations. Thus the user is required to conform to the operational procedures and learn the "language" of the machine. Machines do not provide intuitively understandable interfaces with the consequence that many users of a machine are unable to utilise its complete functionality.

In order to address such interaction problems between people and machines, intelligent socially capable machines must be developed which are able to analyse and understand various natural and temporally coordinated input modalities such as intentions of others which can be inferred from communication (through language, and facial and other gestures such as eye and body behaviours). Through an analysis of the particular context and the fusion of different modalities, faulty and incomplete perceptions may be meaningfully interpreted. (see also Wahlster *et al*., 2001). Psychological experiments are currently exploring how motion can be employed to realise such complex notions as perceived emotional states and deliberative reasoning (Duffy, 2002). These all require an investigation of the role of anthropomorphism in robotics.

4 Anthropomorphism

Anthropomorphism (from the Greek word *anthropos* for man, and *morphe*, form/structure) can be viewed as the tendency to attribute human characteristics to inanimate objects, animals and others with a view to helping us rationalise a situation. It is attributing cognitive or emotional states to something based on observation in order to rationalise another's behaviour in a given social environment. This phenomenon has been exploited from religion in Homer's gods of ancient Greece to recent animation films like "Chicken Run" (Aardman Animations, 2000) or "Antz" (DreamWorks & SKG/PDI, 1998). Accordingly, it has been open to criticism with a first example found in *Fragments* where Xenophanes (c. 560-478BC) wrote a scathing critique of the Homeric texts:

> *"But mortals suppose that the gods are born (as they themselves are), and that they wear man's clothing and have human voice and body. But if cattle or lions had hands, so as to paint with their hands and produce works of art as men do, they would paint their gods and give them bodies in form like their own-horses like horses, cattle like cattle"* (Xenophanes in Lesher, 1992).

Xenophanes' claims articulate an important facet of attributing nonhuman entities with human characteristics. Anthropomorphic activity can complicate simple distinctions between human and nonhuman, while on the other hand, it helps us cope with unfamiliarities and intangibles.

From a psychological perspective, anthropomorphism has not been rigorously studied to any major extent where only a few have seen it as worthy of study in its own right (e.g. Caporael, 1986; Eddy *et al.*, 1993; Tamir & Zohar, 1991), rather than the common view that it is a hindrance to science. Anthropomorphism is considered a hindrance when confounded *with* scientific observation rather than been studied more objectively and taken advantage of. Anthropomorphism *is* a complex and subtle phenomenon (Eddy *et al.*, 1993).

4.1 Use and Misuse of Anthropomorphism

One way of understanding an entity's behaviour is to ascribe mental states to it by simply anthropomorphising it. This involves ascribing human mental characteristics with often little or no reference to the entity's real competences. Eddy *et al.* (1993) suggest that people anthropomorphise animals depending on:

> *"(1) the degree of physical similarity between themselves and the species in question (e.g. primates,) and (2) the degree to which they have formed an attachment bond with a particular animal (e.g. dogs and cats)"*

Caporael (1986) suggests that anthropomorphism is a *"'default schema' applied to non-social objects"*. Eddy points out that familiarity increases the tendency to anthropomorphise (Eddy et al., 1993).

Watt points out that "[a]nthropomorphism is part of what makes us see others as minds rather than as bodies" (Watt, 1998). From a methodological perspective, using such anthropomorphic projection is seen as bad scientific practice, particularly in biology and ethology. Similarly, in cognitive science, Searle (1992) comments:

> *"Prior to Darwin, it was common to anthropomorphise plant behaviour and say such things as the plant turns its leaves towards the sun to aid in its survival. The plant 'wants' to survive and flourish and 'to do so' it follows the sun".*

Searle then tries to provide a more functional explanation:

> *"Plants that turn their leaves towards the sun are more likely to survive than plants that do not".*

Is it possible to remove anthropos from our science when we cannot directly view something from another entity's perspective? Watt points out that *"while we may sympathise with Searle's criticism of anthropomorphism in cognitive science, it may well be endemic*". Eddy *et al.* (1993) note that it is "*almost irresistible*". Krementsov and Todes (1991) 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 won't help. Caporael (1986) 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.

Molloy (2001) highlights that *"[w]hilst particular scientific discourses find the ascription of human characteristics to nonhuman entities interminably fallacious, anthropomorphism also emerges where taxonomic legitimacy of the classification 'human' is under threat"*. Kennedy goes as far as to say that anthropomorphic interpretation '*is a drag on the scientific study of the causal mechanisms*' (Kennedy, 1990). Building social robots forces a new perspective on this. When people are the motivation for social robot research, then people have an influence on how the robot is realised.

An area where anthropomorphism has become a strong issue is in Human Computer Interaction research (Nass & Moon, 2000; Bates, *et al.* 1992; André *et al.*, 1999; Reilly, 1996) with few doubting its strengths. With regard to realising synthetic personalities in artificial intelligence, the assumption has generally been that the creation of even crude computer "personalities" necessarily requires considerable computing power and realistic human-like representations. On the contrary, investigations using simple scripting of text has demonstrated that "even minimal cues can mindlessly evoke a wide range of scripts, with strong attitudinal and behavioural consequences" (Nass & Moon, 2000). The anthropomorphic design of human-machine interfaces is inevitable. The important criterion is to seek a balance between people's expectations and the machines capabilities.

4.2 Two Kinds?

If anthropomorphism refers to the tendency to attribute human characteristics to an entity, two strongly related situations may arise. The first is where we see something and rationalise its behaviour based on humanlike analogies, i.e. "the robot looks lost", "it's thinking". This would correspond to Watt's projective anthropomorphism where we project our interpretations of what is happening onto the entity strictly from an observer perspective: "*external animals and systems become chimerae through the superposition of aspects of the observer*" (Watt, 1998). We draw humanlike assumptions about an entity's behaviour. We may ascribe stronger mentalistic notions than actually exist. It involves rationalising behaviour through mechanisms that we can personally relate to.

The second is the use of anthropomorphic paradigms to augment the functionality and behavioural characteristics of something (both anticipatory and actual) in order that we can relate to and rationalise its actions with *greater ease*. The use of humanlike features for social interaction with people (i.e. Kismet (Breazeal, 2000)) facilitates our social understanding. It is the explicit designing of anthropomorphic features (i.e. two eyes and a mouth) to facilitate the social interaction. These two aspects are not distinct, but rather could be viewed as being two points along the anthropomorphism continuum. This highlights that anthropomorphism is fundamentally observer dependent.

Anthropomorphism can also lead to misunderstanding when the metaphor is either misinterpreted or carried too far. It can lead to an observer who may not be aware of the complexity of a given situation and call the robot "stupid". The observer can similarly attribute behavioural characteristics synonymous with human behaviours to systems that could, in fact, be "trying" to perform something else entirely. A typical example in autonomous mobile robots is a wandering behaviour misconstrued as "the robot is looking for something!"

Watts proposes a notion of "introjective anthropomorphism" in which the "*observer comes to be, in part, a chimera with the observed system*". This "introjective" notion does not constitute a form of anthropomorphism as it involves the modification of the person's behaviour to the observed subject's, i.e. meowing at a cat. Not even if the subject is another person could this be termed anthropomorphic, as anthropomorphism inherently refers to the ascription of humanlike characteristics to nonhuman entities.

4.3 Optimal Anthropomorphism?

Is there a notion of "optimal anthropomorphism"? What is the ideal set of human features that could supplement and augment a robots social functionality? When does anthropomorphism go too far? Using real world robots poses many interesting problems. Currently a robot's physical similarity to a person is only starting to embrace basic humanlike attributes, and predominantly the physical aspect in humanoid research. A robot not embracing the anthropomorphic paradigm in some form is likely to result in a persistent bias against people ascribing mental states to them. Molloy proposes that *"[f]amiliarity, fortunately, offers us a way out of this trap – we can in principle learn to see computers as minds"* (Molloy, 2001).

Research in HCI and robotics has 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 (Hara & Kobayashi, 1995; Breazeal, 2000; Cozzi *et al.*, 2001) often highlighting the importance of making eye contact and incorporating face and eye tracking systems. Interestingly, it can be argued that the most successful implementation of expressive facial features is through more mechanistic and iconic heads such as Anthropos (Duffy, 2002) and Kismet (Brezeal, 2000). Strong human-like facial construction in robotics (Hara & Kobayashi, 1995) has to contend with the minute subtleties in facial expression, a feat by no means trivial. Contrarily, it can be argued that successful highly humanlike facial expression is an issue of resolution.

An important question is how does one manage anthropomorphism. The previous examples demonstrate two methodologies that employ either a visually iconic or a strongly realistic humanlike construction (i.e. with synthetic skin and hair) for facial gestures in order to portray artificial emotion states. The iconic head defines the degree through which anthropomorphism is employed through the robot's construction and functional capabilities and therefore 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.

The question of optimal anthropomorphism is an extremely difficult one, but if we look to try and break it down within a particular context, we may begin to understand its usability and advantages as a conceptual paradigm. It remains to be explored and hints that robot designers must bootstrap it through psychological investigations.

5 Conclusions

The criticism of anthropocentricity is an important issue in cognitive science. Is it appropriate? Should we clarify the distinction between research on *human-like* artificial intelligence and other forms? While anthropomorphism inherently relates to humans, is the phenomenon itself really restricted to our own species? If we refrained from seeking to replicate human intelligence, could we succeed more quickly in achieving a notion of "robot intelligence"? This extreme perspective does highlight one aspect of how we perceive our environment. But there is an obvious counter-argument. Can we realistically decontextualise ourselves from our perceptions? Can we be objective in our view of the environment? Isn't the use of pattern recognition a powerful tool to rationalise for example the behaviour of a robot as it moves about the environment? "It's looking for food!", "it's dancing!". Emergent intelligence takes this observability aspect a stage further. With the implementation of relatively simplistic sensory-motor mappings as found in the Braitenburg vehicles (Braitenberg, 1984) and in emergent intelligence experiments with collectives of simple robots (Cao *et al.*, 1997), it is the observer that has the overall perspective of the system. It is our anthropomorphisation of the robot system that often dictates how we define its behaviour, and consequently what the system is.

The perspectives discussed here have tended towards the pragmatic by viewing the robot as a functional tool employing such humanlike 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 acheive. Even the robot PINO (Yamasaki *et al.*, 2000) 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.

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.

6 References

André, E., Klesen, M., Gebhard, P., Allen, S., Thomas, R., (1999) Integrating models of personality and emotions into lifelike characters, *Proc. workshop on Affect in Interactions – Towards a new Generation of Interfaces, 3rd i3 Annual Conference, Oct., Sienna, Italy*, 136-149.

Bates, J., Loyall, A.B., Reilly, W.S., (1992) An architecture for action, emotion, and social behaviour, CMU-CS-92-144, Carnegie Mellon University.

Braitenburg, V. (1984), *Vehicles - experiments in synthetic psychology*. MIT Press

Breazeal, C. (2000), Sociable Machines: Expressive Social Exchange Between Humans and Robots, *Sc.D. dissertation, Department of Electrical Engineering and Computer Science, MIT.*

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., Williamson, M., (1998) The Cog Project: Building a Humanoid Robot**,** *in C. Nehaniv, ed., Computation for Metaphors, Analogy and Agents,* Vol.1562, *Springer Lecture Notes in Artificial Intelligence*

Cao, Y.U., Fukunaga, A.S., Kahng, A.B. (1997) Cooperative Mobile Robotics: Antecedents and Directions, *Autonomous Robots,* 4, P1-23,

Caporael, L. R. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior*, 2(3):215–234.

Cozzi, A., Flickner, M., Mao, J., Vaithyanathan, S., (2001) A Comparison of Classifiers for Real-Time Eye Detection, *ICANN, P993-999*

Duffy, B.R., (2002) The Anthropos Project: http://www.mle.ie/research/socialrobot/index.html

Duffy, B.R., Collier, R.W., O'Hare, G. M. P., Rooney, C.F.B., O'Donoghue, R.P.S. (1999) Social Robotics: Reality and Virtuality in Agent-Based Robotics, *Bar-Ilan Symposium on the Foundations of Artificial Intelligence: Bridging Theory and Practice (BISFAI), Israel*

Duffy, B. The Social Robot. (2000) *Ph.D. Thesis, Department of Computer Science, University College Dublin*.

Eddy, T.J., Gallup, G.G., Jr. & Povinelli, D.J. (1993). Attribution of cognitive states to animals: Anthropomorphism in comparative perspective. *Journal of Social Issues*, 49, 87-101.

Gallup, G.G., Jr., Marino, L.A. & Eddy, T.J. (1997). Anthropomorphism and the evolution of social intelligence: A comparative approach. In R.W. Mitchell, N.S. Thompson & H.L. Miles (Eds.), *Anthropomorphism, Anecdotes, and Animals*. (pp. 77-91). Albany, NY: State University of New York Press.

Hara, F. Kobayashi, H., (1995) Use of Face Robot for Human-Computer Communication, *Proceedings of International Conference on System, Man and Cybernetics,* p10.

Haraway, D. (1992) Otherworldly Conversations; Terran Topics; Local Terms. *Science as Culture, 3*(14), 64-98.

Iida, F., Ayai, H., Hara, F., (1999) Behaviour learning of a face robot using human natural instruction, *8th International Workshop on Robot and Human Interaction, Sept., Pisa, Italy*

Kennedy, J.S. (1992) *The New Anthropomorphism*. Cambridge Press.

Krementsov, N. L., & Todes, D. P. (1991). On Metaphors, Animals, and Us, *Journal of Social Issues, 47*(3), 67-81.

Lesher, J.H. (1992), *Xenophanes of Colophon: Fragments: A Text and Translation With a Commentary,* Univ of Toronto Pr; ISBN: 0802059902

Nass, C., Moon, Y. (2000) Machines and mindlessness: Social responses to computers, *Journal of Social Issues, 56*(1), 81-103

Nilsson N.J. (1984), Shakey the robot, *SRI A.I. Center Technical Note 323, April 1984*

PINO http://www.symbio.jst.go.jp/~yamasaki/

Rao, A.S., Georgeff, M.P. (1991), Modelling Rational Agents within a BDI Architecture, *Prin. Of Knowl. Rep. & Reas., San Mateo, CA.*

Reilly, W.S.N., Believable social and emotional agents, PhD Thesis, CMU-CS-96-138, Carnegie Mellon University

Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, Massachusetts: MIT Press.

Tamir, P., & Zohar, A. (1991). Anthropomorphism and Teleology in Reasoning about Biological Phenomena, *Science Education, 75*(1), 57-67.

Wahlster, W., Reithinger, N., Blocher, A., (2001) SmartKom: Multimodal Communication with a Life-Like Character, Eurospeech, Aalborg, Denmark.

Watt, S. (1998) Seeing this as people: anthropomorphism and common-sense psychology, *Ph.D Thesis, Knowledge Media Institute, The Open University.*

WE-3RIV: http://www.humanoid.waseda.ac.jp/

Yamasaki F., Miyashita T., Matsui T., Kitano H., (2000) PINO the Humanoid: A basic architecture, *Proc. of The Fourth International Workshop on RoboCup, August 31, Melbourne, Australia*