Emotion Machines: Projective Intelligence and Emotion in Robotics

JOHN BOURKE, BRIAN DUFFY
Anthropos Project
Media Lab Europe
Sugar House Lane, Bellevue, Dublin 8
IRELAND
{johnb, brd}@media.mit.edu http://anthropos.mle.ie

Abstract: - This paper investigates the measurement of social robot performance in order to develop and understand man-machine interaction. It outlines some of the critical points influencing human robot interaction, and details an experiment that demonstrates the willingness of people to treat robots as if they had some human characteristics. It discusses the prevalent factors in the assessment of this performance, and investigates the presence of anthropomorphism in robotics, using the exhibition of motion behaviours by robots to elicit reaction from observers.
Key-Words: - Projective intelligence, anthropomorphism, artificial intelligence, autonomous robots

1 INTRODUCTION

As humans, we tend to use what we know to justify and reason what we see. By the time we have matured into adults, we will have experienced many things in the world. We will use these experiences to provide meaning and logic to new experiences we encounter. This process is a form of pattern matching and results in our ability to deal with unknown situations based on previous experiences that are often associated with more recent definitions of intelligence. When humans interact with machines and other animal species, we tend to ascribe human qualities and traits to these objects. The reason we do this is so we can provide a better understanding of our environment. While this projection of human characteristics onto robots has historically provided considerable problems for the scientific domain, it is nevertheless a fact of human social interaction, and must therefore be considered in any situation in which a human engages in social interaction with a robot.

Measuring performance in robotics has always been difficult due to the differing aims of each individual project, and also the varied platforms on which they are developed. The trend in science has too often been to measure everything, analyse the results, and hopefully come up with a figure that demonstrates some level of achievement or gain. This can be a very restrictive process, as limited information can be recorded fully with numbers, and can result in the discarding of much information. Often a method that can record more descriptive results would be preferred. A question to be answered is what exactly are we measuring? In many cases where the objectives of a robot are simple, measuring the success or failure of various attempts can be relatively simple. But when a robot has a complex task such as social interaction, where do we start, and what do we measure? Who decides if an interaction is a success? What do we measure the robot against? When humans are undertaking the measuring, the obvious benchmark to which the robot is compared is the human being and unless we develop a fully synthetic human being, all robots will be classed as a failure when compared to a human. The following sections discuss the notion of artificial intelligence as applied to robots and the implications of social interaction with robots.

2 WHAT IS ROBOT INTELLIGENCE?

Minsky's 1968 definition states that "…artificial intelligence is the science of making machines do things that would require intelligence if done by people". This definition involves mankind immediately. Machine intelligence is inherently associated with humans and it is man who measures this intelligence. Man also creates the standards against which we measure the intelligence of robots, and in the end it is Man who going to decide if a machine is intelligent. We can therefore see that it is impossible to take the human out of the system, as it is he who sets benchmarks for intelligence, and determines if a robot is actually classed as intelligent.

In its simplest form, an intelligent agent is one that does intelligent things. So how do we know that a robot has done an intelligent thing? Presumably, the robot would firstly need to have the intention to achieve some goal, and then actually fulfil that goal. Therefore, if we had an agent, whose goal was to navigate from point A to point B on a flat surface, and the robot navigated to point B with no collisions, and in an efficient manner, we could say that the robot had completed the task correctly (similar to a calculator adding 1+1 resulting in 2). However, because it is a human that will decide if a machine is intelligent, a particular individual could observe a robot carrying out this task, and think that the robot was indeed intelligent, and that it had made its movements based on an inherent desire to navigate the surface (Searle discusses intentionality, its implications and misuse in "*Intentionality: An Essay in the Philosophy of Mind*" [1])

Because the real world is complex, there are many unknowns for a robot to encounter; completing a task correctly in a controlled environment is a simpler task than if a machine was placed in the real world. The robot Shakey [2] illustrated this and heralded the rethink of the classical AI stance in seeking to realise a robot that displays a degree of intelligence. It is impossible to

anticipate all the difficulties that a machine will encounter, and consequently it is impossible for a machine to be ready to cope with everything it will experience in the real world [3].

In a world where there is growing work in the area of social robotics [4][5][6], the means to assess the performance of social robots are becoming more difficult to design. These robots exist in a complex, dynamic environment, and the measurement of success or failure is becoming increasingly more difficult. One of the problems is that as we aim to embed robots into our physical and social space, the human inputs on the 'intelligence' of the robots, as it is a space designed by humans to suit human needs. Consequently, a robot would have to adapt its ways to accommodate humans (and to an extent, vice-versa). The performance assessment of these robots depends on an individual human, and it is this subjectivity that causes difficulty when defining a robot's 'intelligence'. One human might think a robot intelligent, because it can carry out whatever tasks this human requires, while another human might not think a robot too intelligent, because he wants it to carry out a different task and it is not designed to do so. The robot can be perceived as overly stupid when it fails in what appears a simple task, even though it may be capable of more complex behaviours. This leads to the notion of *perceived intelligence*.

3 PERCEIVED

The perceived intelligence [7] of the agent is closely linked to the agent's goals within a particular environment. A robot, whose actions seem very intelligent in one environment, might not seem so intelligent when the environment is changed. The same is true in real-life. For instance, a doctor who seems very intelligent in his own environment might not perform so well when placed in a farmyard and asked to function successfully there. So in the case of any measurement of intelligence, context is everything. The intelligence of a robot depends on the environment in which the robot exists.

But does a robot have to be *inherently* intelligent, in order to be *perceived* as intelligent? As mentioned, a robot whose task is to navigate a controlled environment, might have a very simple control program, with some sensors and motors, to provide the means necessary for successful navigation. To an observer, this agent might indeed look intelligent.

Having realised this, we then have another question to answer. If a person/robot looks intelligent to observers, is that enough? Humans make many decisions on appearance, so if a robot appears to be intelligent in a situation, does the human observer then think that this robot has some form of inherent intelligence? For humans to think that a machine is intelligent, it is likely that machine would have to appear intelligent. And, if this is the case, is it enough to appear intelligent without the backup of real inherent intelligence? Can this illusion of intelligence be maintained over time? While a human might be fooled for a short period, it is likely that the appearance of intelligence would be short-lived if the robot takes an unintelligent action.

4 MEASURING PERFORMANCE

Metrics in Science has always tended towards using facts, figures, and data to justify findings. Artificial Intelligence (AI) and robotics, have maintained this status quo by generally neglecting to consider other methods by which the performance of a robot can be measured. The tendency in science has been to measure everything using numbers and standards, while in social sciences such as psychology; a more subjective analysis is carried out, but is still reduced to a compilation of facts and figures to support conclusions and theories. This quantification of results can cause much information to be lost, as often it is impossible to record the rich detail of experience in a summary of numbers.

To date, these empirical methods have generally failed to realise a constructive generic assessment of robot performance. The first reason for this is embodiment. Early philosophers such as Descartes, who studied humans as if they were mechanical things, maintained that the mind was separate from the body. A 'thinking thing' would only require a mind, and although it may also have a body, it was not necessary. The body could be used for sensory perception, but it was not part of the 'spirit' of a human. Steels [8] argues that for a machine to be intelligent, it must have a notion of embodiment and this is backed up by Brooks who maintains that embodiment is vital to the development of artificial intelligence [9]. There is no doubt that embodiment provides a physical and even social context for robot behaviour, and without physical existence in the real world, a robot's experience of the world would be underdeveloped and highly restricted. For a robot to better understand the world, it needs to have a physical body to provide a fuller experience and this brings with it the complex issue of assessing a robots performance in a world that has few constraints and difficulties in anticipating difficulties that a robot might encounter.

Another reason for the failure of empirical methods of robot assessment is that, while we can measure and record performance figures, there are no standardised benchmarks against which these can be compared. Consequently, a more abstract measurement of a robot's performance is needed, encompassing all tools and indicators available in order to get a fuller, more accurate estimation.

In most cases, the success of a robot depends on its design goals. For instance, if a robot is designed to navigate a path from one end of a room to another, then a good measurement of its success, would be to measure the time this navigation takes to complete, and to count the number of times the robot bumps into obstacles on the way. The faster a robot completes its task, with the lowest number of collisions, the more successful it is. Very obvious, you might think, but this assessment depended on this particular robots simple aim. The assessment of a robot's performance is abstracted away from the

underlying mechanics, and the behavioural mechanisms that are employed to achieve a goal.

Alternatively, an observer of various robots' attempts to navigate this room might draw much the same conclusions as collected empirical data purely by looking at the robots' attempts at getting from start to finish, and judging which one had made the best effort at the completion of the task. But the problem is how do we decontextualise the observer? How do we isolate the observed results from observer- dependent perspectives? Programming a robot yields a similar problem. We design a robot's behaviour from the perspective of how we would rationalise a problem – not necessarily how a robot would. Humans are at all times inputting to the robot's design, and it becomes impossible to completely detach humans from a robot's design. We might think that that the solution for this problem would lie somewhere within the robot learning area of AI, but even the original learning algorithms are all designed by humans, returning observer dependency ('contamination') to the loop.

As technological advances are made in robotics hardware, the design goals of various robots become more ambitious and complex. As robots become more useful to humans, and as their behaviour becomes more complex, their use as tools will become more common necessitating an observer-based assessment of the performance. The robots will be working for us so it will be a human who will decide on their performance. To illustrate, perhaps a simple room navigation task might be combined with collecting particular objects on the way, either randomly or in a specified order. Obviously a bit more work is required of the robot, but once again, the measurement of success can me made by collecting data, with subsequent analysis, or by simple observation and judgement.

As tasks become more and more complicated, with different subtasks, the assessment of which design is better than which, becomes much more difficult to decide. Similarly, the behaviour generation algorithms become arbitrary. If the robot performs a task successfully, the particular motion control and digital signal processing algorithms they employ also become arbitrary. The RoboCup Tournament (http://www.robocup.org/) is an example of this thinking. In this tournament, the objective is to score more goals than your opponent, and the method or programs employed to achieve this task are not important. It is the success and completion of the task that are important. This is a key change of perspective in performance analysis, which this paper aims to highlight. Often the use of data acquisition, to decide an outcome, becomes too complex, and the winner can depend on which areas of a robot's deign parameters and goals, a particular human observer places the highest priority.

Because the measurement of robot performance is so specific to a particular instance, academic works that aim to provide a general robot assessment mechanism are infrequent. There is no universal standard for benchmarking because there are many different robots, with different goals, being developed on different platforms.

5 ANTHROPOMORPHISM

The appearance of a robot becomes an issue when assessing its performance. When humans interact with machines and other animal species, we tend to ascribe human qualities and traits to these objects. When humans view the actions of animals or machines, they rationalise the behaviour through anthropomorphism. This process begins in childhood when we talk to our teddy bears and become attached to a particular blanket for security. We carry this trait through to adulthood, when we treat pets and objects such as cars as if they have human qualities. With the advent of technology, humans now become attached to robot dogs, and the Tamagotchi is an example of how we become so connected to them that we have to nurture and look after them as if they were human.

This phenomenon has been around for some time and as Krementsov and Todes [10] stated, *"the long history of anthropomorphic metaphors … may testify to their inevitability"*. Science, on the other hand, has tended to try to isolate this factor from its progress, view it as a factor to be removed from experimentation, and only tested in its own right. But is this the right attitude to have? Should we not employ this human characteristic in any effort at social interaction? The willingness of humans to anthropomorphise inanimate objects becomes a useful tool that can allow robot designers to more easily provide interaction between humans and robots. Can it be harnessed to allow mechanistic devices become more intelligent in the eyes of humans?

This 'observer judgement' method of performance assessment brings with it some interesting points. To achieve a fair result, total impartiality must be maintained at all times, and is maintained in the Turing Test by using typewritten word only. A point worth noting here is that the use of type written word for interaction between the two 'systems' in the Turing test also highly constrains the interaction and inherently influences the intelligence assessment. This impartiality is often impossible to achieve, because as humans mature into adults, each has gone through an infinite number of experiences of happiness, sadness and troubles. No one person has exactly the same experiences as another, and therefore it is impossible to regulate a performance test (by human observation) to a point where the observers' life experiences have no effect on their judgement of others, and consequently their opinion of relative success or failure. A crucial factor in this is that it is humans who design these robots, making a totally unbiased system impossible to achieve from the very outset.

However, we can judiciously use anthropomorphism to enhance human opinion of a robot's performance. We allow our own experience bias our interpretation of what we see. If we accept that it is difficult for humans to objectively judge robots (and other humans) and by doing so adopt the view that anthropomorphism is inevitable, we should then use this trait as best we can, in our high-level human computer interactions. And so, we come back to perceived intelligence. If a human is willing to employ anthropomorphism (whether consciously or not) to rationalise a machine's actions, then why not use this to our gain? Some would argue that the use of

anthropomorphism in human-computer interaction allows unpredictable consequences to be embedded into design, but although these arguments are valid in theory, they become idealistic when designing social robots [11]. Although anthropomorphism is subjective, there will undoubtedly be indicators that prompt human observers to project the right characteristics onto a machine.

Another issue to be addressed in the measurement of robot intelligence is what do we measure against? Undoubtedly the highest benchmark standard of a social robot has always been the Human Being. Indeed, Webster's English Dictionary defines the word robot as "*any manlike mechanical being, by any mechanical device operated automatically….*". The word robot itself provides associations between machines and man. There have been many works that attempt to artificially imitate the human form [12][13][14]. The inherent fear of robots (and other unknowns) by people is indicated by early films and plays such as Capek's "Rossum's Universal Robots". Even Asimov's laws of robotics, produced in the 1950's are designed to allay the fears of people that robots will become too strong and self-sufficient, and will 'take over the world' [15]. This close mimicking of humans, and hiding of the mechanical interface underneath, can allow more comfortable social acceptance by humans themselves – up to a point. The fear of artificial beings becoming uncontrollable heightens as robots become very close to humans in appearance. In the early 1980s, Mori illustrated this phenomenon with "The Uncanny Valley". He argued that as robot faces became more like human faces, there would be a range where the robot face becomes 'uncanny' and strange to a human observer.

A pitfall of aiming to closely imitate the human form is that the best attempts at this imitation are flawed. The human observer will find it easy to point out these flaws. As a robot's appearance becomes closer to that of a human, we leave ourselves open to an overestimation of the robot's intelligence. Therefore, a robot that has a low degree of similarity to humans could be accepted as an intelligent machine, but if the robot becomes too close to human in appearance, it can then become a stupid human. There is a constant need for robotics engineers to thread this fine line when developing a synthetic manlike being.

6 ROBOT EXPERIMENTS

The presence of anthropomorphism in man-robot interaction is researched using Khepera robots that exhibit various motion behaviours to elicit a reaction and interpretation from human observers. These behaviours, based on the emotional characteristics associated with Disney's seven dwarfs, were recorded on film and placed on a website where observers could record their observations.

People recorded their own thoughts and feelings about the robot behaviours. It enabled a comparison of their reactions to be made, and documented how the robots' motion behaviours affected the observers' impressions of them.

In a human, there are obvious indicators to the emotional state of a person such as facial expression and speech. Because these robots have no such external indicators, any projection of human qualities such as intelligence or emotion had to be hinted at by the movement of the robots round the environment. This presented difficulties as this restriction places constraints on how one might distinguish one dwarf from the other purely by movement. The particular movements of each robot were based on an interpretation of how each dwarf might move. The object of this exercise is not for people to instantly recognise the robots as the Seven Dwarves, and so declare the programmed behaviours 'a success', but rather to allow people to project their own feelings and emotions onto the robots. The movement of each robot would allow the observers interpret the motion behaviours in a way that allows the ascription of interesting characteristics to the robots.

A similar selection of recorded behaviours was made using Kheperas that had some of their underlying hardware hidden by 'dressing' them with some clothes. The purpose of this was to investigate if the tendency to anthropomorphise the robots was affected by the concealment of some of the mechanics.

Some of the people who took part in this experiment concentrated their efforts on describing exactly what moves the robot was taking. Efforts were made to explain the behaviours from a purely technical aspect, with 'searching' and 'learning' very common words used in the replies. However, it is useful to note that people seemed to see past the mechanics of the robots, and began to describe them in terms usually reserved for humans. Unless a fully functional synthetic human is invented, any robots that will work with humans in the home or workplace will seem somewhat mechanical, either in their appearance, communication, movement etc. Therefore, it will be very difficult to create an embodied robot that can fool or trick a human into thinking it is human, especially over longer periods of time. Consequently, it is of practical use to realise that humans can relate to mechanical devices as if they had somewhat human characteristics.

7 RESULTS

The results have confirmed a number of things. Firstly, people are quite willing to see past the mechanistic exterior of robots, and treat them as if they possessed human qualities. This anthropomorphism of these robots could be tactfully employed to enhance any interaction between humans and robots. The fact that a common used word used in the descriptions was 'learning' combined with the lack of learning algorithms in any of the programs, showed that people can be fooled into ascribing characteristics such as intelligence and other traits onto objects that don't necessarily possess them. This projective intelligence shows that people can read more into a situation then actually exists, and is an example of how the appearance of an object can affect people's reactions to it. This fooling of humans is not a malicious or wicked act to be discounted from scientific tests, but is a natural aspect of human life, and as such, can be used as a tool to aid robot designers in their aim to build socially

capable robots, and so should be integrated into any testing of human-robot interaction. Also, given that social robots are designed by humans to exist in a world adapted to humans, and that the assessment of a robot's performance is measured by humans, it is impossible to completely remove human subjectivity from this measurement.

Some of the people who took part in this experiment concentrated their efforts on describing exactly what moves the robot was undertaking. Efforts were made to explain the behaviours from a purely technical aspect, with 'searching' and 'learning' very common words used in the replies. However, it is useful to note that people seemed to see past the mechanics of the robots, and began to describe them in terms usually reserved for humans. Unless a fully functional synthetic human is invented, any robots that will work with humans in the home or workplace will seem somewhat mechanical, either in their appearance, communication, movement etc. Therefore, it will be very difficult to create an embodied robot that can fool or trick a human into thinking it is human, especially over longer periods of time. Consequently, it is of practical use to realise that humans can relate to mechanical devices as if they had somewhat human characteristics from a design perspective whilst also maintaining the reference of the robot being a machine with its own set of capabilities.

The dressed Kheperas considerably changed the overall tendency of people to describe in factual detail what they saw. Many people referred to the 'big eyes' and the 'funny hat'. Replies to this elicited some interesting (and amusing) comments such as "reminded me of 'Furbys'. Looked very cute and noticed the lights of (apparent) reaction when Black robot approached white and red robot. Interaction was cute appeared to be snuggling." Another mentioned "the blinking thingy underneath its 'clothes' reminds me of the Teletubbies' hearts and thought the robots were 'playful'. The fact that these robots were 'dressed' may have softened them in the eyes of the observers, and it does seem to have changed their views substantially.

There was less emphasis on the actual movements within the video clips, and more on what typical human characteristics they looked or acted like. For instance quite a few people mentioned that Grumpy looked like a 'sentry' after viewing the first web page but in replies to the second page, many people thought Grumpy's actions made him looked 'forlorn' and 'frustrated'.

8 CONCLUSIONS

This particular study attempted to demonstrate the effect that motion behaviours have on the projection of human characteristics by human observers onto inanimate objects. Previous research has demonstrated that these inanimate objects can range from a comfort blanket for a child to a state of the art Tamagotchi robot for an adult. This willingness of humans to act in this way is an inherent trait of the human race, as it helps us to use what we have previously learnt in life to understand new experiences. Consequently, when we experience robots, we draw on our previous experiences, from films and stories and life in general, to compare this new object with something with which we are familiar. As the context here is social interaction, the familiar object is our fellow man. The results of this research show that anthropomorphism is a prevalent factor when humans view the behaviours of robots. While this is a small scale and constrained demonstration of its existence and inevitability, it does provide a basis for a more wide-ranging study into how we can use inherent human traits to augment the interaction between humans and robots.

If anthropomorphism is impossible to completely eradicate from robot design, we should then embrace it, and carefully employ it to help people relate to and communicate with robots. Judicious anthropomorphism should facilitate rather than complicate.

References:

[1] Searle, *Intentionality:An Essay in the Philosophy of Mind*, Cambridge University Press, New York 1983.

[2] Nilson, Shakey the Robot, Technical Note 323, SRI A.I. Center, 1984.

[3] Simon, Administrative Behavior:A Study of Decision-making Processes in Administrative Organization (2nd ed.). New York: Macmillan, 1957

[4] Breazeal C., *Designing Sociable Robots*, MIT Press, 2002.

[5] Duffy B., *The Social Robot*, Ph.D. Thesis, Department of Computer Science, University College Dublin, 2000.

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

[7] Duffy B., *Anthropomorphism and the Social Robot*, Robotics and Autonomous Systems journal paper, 2003.

[8] Steels L., *Building Agents with Autonomous Behaviour Systems, The "artificial life" route to "artificial intelligence"*. Building situated embodied agents. Lawrence Erlbaum Associated, New Haven, 1994.

[9] Brooks R.A.,*Integrated Systems Based in Behaviours*, SIGART Bulletin 2(4),1991,p46-50.

[10] Kremenstov N., Todes D., *On Metaphors, Animals and Us*, Journal of Social Issues 47(30), 1991, p.67-81.

[11] Shneiderman, B. *A nonanthropomorphic style guide: Overcoming the humpty-dumpty syndrome*. The Computing Teacher, October, 1988.

[12] Master Lee: YFX Studio, http://www.yfxstudio.com/human.htm

[13] Saya: Kobayashi Lab, koba0005.me.kagu.sut.ac.jp/newsinfo.html

[14] Roberta: Science University of Tokyo, hafu0103.me.kagu.sut.ac.jp/haralab/

[15] Asimov I. *I, Robot*, Doubleday, 1961.