2005

Future Reasoning Machines: Mind and Body

Brian Duffy
Media Lab Europe

Gregory O'Hare
University College Dublin

John Bradley
University College Dublin

Bianca Schoen-Phelan
Dublin Institute of Technology, bianca.phelan@dit.ie

FUTURE REASONING MACHINES: MIND & BODY

Brian R. Duffy[1], Gregory M.P. O'Hare[2],
John F. Bradley[2], Alan N. Martin[2], Bianca Schoen[2]

[1] Media Lab Europe, Sugar House Lane, Bellevue, Dublin 8, Ireland
brd@medialabeurope.org

[2] Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland
{gregory.ohare, john.bradley, alan.martin, bianca.schoen}@ucd.ie

ABSTRACT

In investing energy in developing reasoning machines of the future, one must abstract away from the specific solutions to specific problems and ask *what* are the fundamental research questions that should be addressed.

This paper aims to revisit some fundamental perspectives and promote new approaches to reasoning machines and their associated form and function. Core aspects are discussed, namely the one-mind-many-bodies metaphor as introduced in the Agent Chameleon work. Within this metaphor the agent's embodiment form may take many guises with the artificial mind or *agent* potentially exhibiting a nomadic existence opportunistically migrating between a myriad of instantiated embodiments.

We animate these concepts with reference to two case studies, illustrating how a machine can have fundamentally different capabilities than a human which allows us to exploit, rather than be constrained, by these important differences.

Keywords: Cybernetics, human machine interaction, robotics, agent systems, anthropomorphism, philosophy of cybernetics

1. INTRODUCTION

Intelligent systems research has undertaken an arduous and evolving path over many decades, all the while delivering, in approximately equal numbers, solutions to many problems whilst also identifying further as yet unsolved problems. Core principles from many disciplines have influenced our perspectives on a system's function and form, none more so than the *one-mind-one-body* debate found in biological entities. The age old notion of embodiment (the strong provision of context within the system), whether physical (Brooks 1991, Steels 2000) or social (Duffy 2000, Duffy & Joue 2001), has been an important development in artificial intelligence research and robotics. It has focused the control strategies employed on the robot's environmental contexts. While fundamentally important and necessary, the continuing focus on the narrow frame of reference of one-mind-one-body should be developed further and new paradigms investigated, which this work aims to address.

Inevitably, our sources of inspiration come from what exists around us. Consequently significant research energies have been invested in such projects as trying to realise a human-like robot, a system that clearly encapsulates the one-mind-one-body concept. But to what extent should a machine's reference be sourced from such biological references as ourselves? Is the human-based approach to a singular mind-body paradigm the only tangible option? Indeed how ought we to manage our perceptions and interpretations of artificial entities that extend beyond this paradigm?

René Descartes, referred to as the father of cybernetics due to his study of the human body as a machine, popularized the age-old thesis that mind and body are distinct from each other (Descartes, 1637). He argues that even though he

may have a body, his true identity is that of a thinking thing alone and, indeed, his mind could exist without his body. He argues that humans are spirits, which occupy a mechanical body, and that the essential attributes of humans are exclusively attributes of the spirit (such as thinking, willing and conceiving), which do not involve the body at all. While this has been considerably debated in the field of AI and robotics over recent decades, it has become generally accepted that embodiment is key to the development of AI in machines. But what if we wanted to build artificial systems that extend *beyond* this traditional paradigm? This paper draws on Descartes notion of "spirit" and extends this to a one-mind-many-bodies paradigm.

The following sections provide firstly, a background discussion of the artificial mind and the issues surrounding artificial intelligence and robotics with regard to the traditional paradigm of one-mind-one-body. Following this, section 4 takes a step away from existing approaches in the context of the reasoning machine and looks at how a "spirit" can be embraced in artificial systems with the added dimension of being able to change and "possess" different bodies. Section 5 looks at some of the core features of artificial systems and argues how the function and form of the machine are inextricably linked, but which are still subject to observer-based dependencies. Finally, section 6 presents some fundamental tenants that underpin the next generation of reasoning machines.

2. THE ARTIFICIAL MIND

The relationship of the mind and body has been a psychological and philosophical problem for many years. From both philosophical and scientific theories, the mind-body relationship can be divided into two main categories: monistic and dualistic. Firstly, monistic theories suggest that mind and body are not independent of one another. Behaviourists (including the likes of Aristotle, Hobbs and Hegel) hypothesised that mind was nothing more than a function of the body. Idealists, like Berkeley, Leibniz and Schopenhauer suggested that the body was just a mental representation. Spinoza proposed a theory of double-aspectism which postulates that mind and body are distinguishable but not inseparable.

Secondly, dualistic theories are of the view that mind is seen as distinct from the body and not made up of any physical substance. Some popular dualists include Descartes, Locke and James who belonged to a branch of dualism known as interactionism. Descartes, as an *interactive dualist*, believed that there was a distinction between the human mind (or soul) and the physical body, describing the mind as "*[a] thing or substance whose whole essence or nature was only to think … has no need of space nor of any material thing or body. … This mind … is entirely distinct from the body*" (Descartes, 1993). He claims that a body without a soul would be an automaton, responding to external stimuli, while a soul without a body would have consciousness but only of innate ideas, lacking any sensory impressions (Francher, 1979). Interactionists believed that, although mind and body were of a very different nature, it was the interaction between the two that produced many aspects of human nature.

The prevailing view in cognitive science today is that the human mind consists of distinct faculties dedicated to a range of cognitive tasks: linguistic, social, practical, theoretical, abstract, spatial and emotional. Mental processes in humans are generally viewed as not being solely internally-represented symbol-manipulating algorithms, and thus the notion of a robot having a *mind* using the human mind as the frame of reference becomes an issue. Arguing against artificial system having a mind is similar to discussing whether the system simply operates at a level of syntax without semantics (i.e. a computer *acting* the role). Programs operating on a machine can be seen to be semantically blind, merely mimicking the grasp of meaning according to both the rule set employed and the data received. It does not *understand* the information; it merely has a methodology capable of dealing with it, a form of mapping between input and output. Searle (1980) animates this stance in his Chinese Room Argument.

This paper argues against applying the term *mind* with respect to machines and makes the distinction between mind and the *artificial mind*. The term artificial mind refers to an artificial entity's reasoning mechanism, independent of particular implementation technologies which have been developed for its interaction with both its physical and social environments. This undoubtedly includes our worlds and therefore its interaction with us. Machines with minds may arguably not exist, but the importance of AI is that machines with artificial minds can exist because as humans we tend to interpret the artificial entity according to our frame of reference. That is, we basically anthropomorphise and adopt the intentional stance (Dennett, 1987)[1].

[1] Consciousness is not discussed in this work although the ideas presented undoubtedly promote this discussion.

3. ROBOTICS AND AI

A predominant theme within AI research is to focus on the development of functional components and solutions to narrow problems, with limited abstraction and consideration of the broader objectives of AI.

Artificial intelligence was initially interpreted as an attempt to prove the *Physical-Symbol System Hypothesis* where "*formal symbol manipulation is both a necessary and sufficient mechanism for general intelligent behaviour*" (Simon, 1957). Efforts to solve the AI problem that follow this hypothesis are now termed the *classical AI* approach. Simon maintained that the human cognitive system is basically a serial device.

When results were subjected to human interpretation, classical AI provided a rich source of control ideas. Problems arose when these control paradigms were applied to robotics, and in particular the control of mobile robots. The original theory that robots would simply provide the sensors and actuators for an artificial brain, when constructed, became seriously flawed.

The robot Shakey (Nilsson, 1984) provided a useful calibration for classical AI and its original idea of developing some form of artificial mind (effectively an artificial reasoning mechanism). While focused strategies to specific solutions are essential, they generally merely provide the mechanisms upon which more complete systems can be constructed. By reviewing achievements and failures to date within robotics and AI, an insight is acquired into the continued relevance and attainability of the grand challenge.

Problems arose with real-time performance and stability through, for example, sensor noise and demands of maintaining representational model validity. More elaborate models necessitated ever increasing computational effort that often proved too cumbersome and not sufficiently responsive for real-world applications. It became apparent that understanding *system-environment* interaction was fundamental in achieving robust control for autonomous robots existing within a physical world.

This classical approach viewed mind as distinct from body and took a non-interactive dualistic approach. Early research in the field of Artificial Intelligence worked on developing artificial minds that were effectively disembodied with minimal interaction with any world (real or otherwise). However, this has a fundamental flaw, in that "*a program integrated in a computer with no visible appearance nor autonomous physical interaction with the real world has a more difficult time to be viewed as intelligent, whatever the power of its problem solving and the sophistication of its knowledge*"(Steels, 2000).

The inability of such classical AI systems to handle unconstrained interaction with the real world led to a search for new control architectures for autonomous robots. Recent research into embodiment, sociality and emotions are now approaching the problem from a new angle. This *New AI* has assumed a stance similar to double-aspectism. While mind and body are viewed by some as distinguished separate components they are not necessarily inseparable. A series of provocative papers by Brooks (1986, 1990, 1991), argued that real world autonomous systems or embodied systems must be studied in dealing with the problems posed by classical approaches. While not a new concept, Brooks' popularisation of the reactive approach served as a useful catalyst in looking for more embodied approaches to artificial cognition. Issues in real-time processing became very real, for example if the robot could not cope and it crashed into something. Only by direct interaction could the robot gain an understanding of the environment.

For either of the deliberative or reactive approaches, a robot requires a control architecture. This architecture determines how behaviour is generated based on signals from sensors and invoking motor responses. Reactive approaches which aggregate large numbers of simplistic non representational reasoners have led to *emergent "intelligent" behaviour* (Braintenburg, 1984; Fukuda, 1989; Kube, 1993; Lucarini, 1993). While founded in embodied robotics, these have not proved sufficient in order to achieve complex goals and suffer from issues of repeatability and the absence of a strong theoretical model. In contrast, deliberative architectures have displayed reflective reasoning capabilities but may lack the responsiveness and robustness demanded by real world applications.

Thagard defines a current central hypothesis of cognitive science, the Computational-Representational Understanding of Mind (CRUM): "*Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures*" (Thagard, 1996). While there is much speculation regarding the validity of this statement, he continues by stating that this central hypothesis is general enough to encompass the current theories in cognitive science including connectionism. Interestingly, while strong embodiment, as discussed in (Brooks, 1986; Thagard, 1996; Duffy, 2000), continues to prevail as a necessary criteria for achieving stronger notions of intelligence in artificial systems, the fundamental mechanisms used in trying to build strongly embodied systems today are inherently symbolic in nature. The process control is achieved via symbolic computers. So, can embodied artificial cognitive processes be really achieved to the extent required to realise a strong notion of intelligence when our references for intelligence are based on natural systems? It's like trying to make machines into natural entities, or inversely, to reduce natural systems to machines, an issue regularly discussed in AI.

Two diametrically contrasting issues arise:

(1) If the reference for intelligence and the barometer for gauging degrees of intelligence is that of the human then anything less than a human in all capacities is unsuccessful.

(2) If the qualifier *artificial* is emphasised, then the process of comparison becomes more of an analogy, *with limitations*.

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 how people think could be revealed through an understanding of the computational processes that govern brain characteristics and function. This would also provide an insight into how one may realise an artificially created intelligent system with emotions and consciousness.

In contrast, advocates of *weak AI* believe that human intelligence can only be simulated. An artificial system may 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.

Having briefly reviewed the AI and robotics journey to date, we wish to pause and consider the function of the reasoning machine, and the associated possibilities in terms of its divergent forms. In clearly distinguishing between artificial and biological systems at a control level, this inherently draws a distinction between their capabilities. If the system is a machine, this effectively centres its functionality on its mechanistic construction, which can be to a designer's advantage.

4. FUNCTION OF THE MACHINE

What is the function of the ultimate reasoning machine? With little references, our ability to invent something beyond the capabilities of what we see around us can become difficult. Could we even understand something so different, let alone invent it?

Robot success to date is based upon an ability to determine those tasks for which the robot is particularly apt. Examples include assembly, repetitive pick and place, hazardous substance manipulation, welding and spray painting. Their role as a tool is clear and their function relies and exploits the properties of machines. Given that machines have a fundamentally different capability set, to constrain it to our capabilities is simply inappropriate. The issue becomes what *could* it do that exploits its inherent functionality? At a very basic level, are the human frames of reference of one-mind-one-body still valid in developing a reasoning machine's functional capabilities? The following subsections challenge the prevailing concept in the field of artificial intelligence of one-mind-one-body and exemplify the principle of this paper which is to not become limited in one's design and development of artificial systems.

4.1 Free the Mind

Like the rationalist tradition in philosophy (Descartes, Leibniz, Kant, Husserl) AI research holds that the mind is fundamentally rational, representational and rule-governed. Because of this, modern philosophers like Dreyfus (1972) argue that AI research will fail because it falls prey to precisely the same issues that were directed against the rationalist tradition in philosophy. Furthermore, an animal mind is an aggregation of a vast number of highly parallel, asynchronous, analogue processes. In contrast, artificial intelligence to date is based on digital devices that in most cases can only give the illusion of parallel, asynchronous behaviour. By their very nature, such devices can only ever give an approximate illusion of (artificial) mind. So, maybe traditional philosophies of mind cannot be applied directly to digital entities but rather should only be used as analogy. Such entities, whether a real robot or a virtual avatar for example, can be viewed as virtual entities where it is impossible for that entity to inherently know for certain whether their instantiated platform *is* a robot or an avatar. It's a control program run on a CPU. As current AI technology is based on digital devices, and as such all input/output and processing in an AI mind is through digital means, they do not necessarily require a fixed platform, just as long as the platforms on which they are instantiated support these computational entities.

It is important to note that this does not necessarily undermine the embodiment debate but rather embraces the context of the artificial entity existing through its body instantiations in a physical and social environment. While being physically and socially grounded is fundamental, the added dimension presented here is that the form of its body can change. It would be incorrect to base arguments against the one-mind-many-bodies idea on embodiment arguments that are grounded in natural systems, i.e. that in order to artificially develop a system that displays intelligence, one must achieve the degree of system-environment integration found in natural systems. An autopoetic system is very

much distinct from an allopoietic system (Sharkey & Ziemke, 2000). The aim here is more to develop *reasoning* machines rather than humanlike intelligent machines. Consequently, in using the prefix *artificial*, this new form of intelligence should exploit its inherent differences to humans rather than be artificially constrained by it.

While strong embodiment prevails as a grounding in aiming to achieve an artificial notion of intelligence in its true form, Descartes' original proposal that mind is distinct from body rises again. By embracing functionality provided by mechanistic solutions ("telepathy" through wireless communication), the long standing notion of one-mind-one-body is challenged.

When applied to Artificial Intelligence, the general Cartesian model suggests that an agent has a distinct mind and body, but yet the mind relies upon the body and the body upon the mind in order for them both to operate successfully. It should be noted that distinctions between the physical instantiation of a body rather than using a virtual representation are often highlighted. It is argued that such physicality is viewed as a requirement for intelligent behaviour where its reference is our real and complex world. This is a different debate and the view adopted here is that the physical context is always present, whether the interaction is through robot actuators or VR interaction devices with users in the physical environment.

Within the context of the one-mind-many-bodies metaphor, an agent's actions can be significantly enhanced:

- Firstly, there is no restriction upon the agent's embodiment form should take. Numerous guises may be adopted, for example that of a physical robot, or an avatar in virtual reality or a small 2D animated gif suitable for display on a PDA.
- Secondly, the artificial mind or *agent* could migrate between a myriad of instantiated embodiments akin to a ghost moving and possessing different bodies. The behaviour of the agent will be dictated not only by the agent's goals but also by the embodiment that the agent has adopted.

The choice of embodiment must not only empower the agent, but maintain the agent's identity in the eyes of the user. This can be achieved in a number of different ways, including:

- Preserving key referential characteristics across the different instantiations, for example the agent's colour scheme or eyes.
- Using transitions that maintain the presence of the agent throughout the transformation process.

The agent is thus unconstrained to any particular environment, physical or virtual. Opportunistic migration both between and within different environments, physical and virtual, should exploit the functional capabilities within each.

One important platform demonstrating this de-restriction from our frame of reference is an agent within a 3D virtual environment. Virtual Environments (VE) or Virtual Reality (VR) have a number of distinct advantages over other forms of real world instantiations (such as robotics). Within VR, the rules of the real world, for instance gravity, need not necessarily apply. The agent form is likewise unconstrained, as it is capable of mutation in order to suit the task at hand, it may even choose to abandon one form of embodiment and adopt an entirely new one.

The objective is therefore to augment the artificial entity's functionality beyond our own frame of reference. The following case study demonstrates the one-mind-many-bodies metaphor.

4.2 CASE STUDY: Agent Chameleons

The Agent Chameleons Project (Duffy et al, 2003; O'Hare et al, 2003) strives to develop digital artificial minds that can seamlessly travel between and within physical and digital information spaces. Three key attributes of migration, mutation and evolution underpin this concept, and can be invoked in response to environmental change, ensuring the survival and longevity of the agent.

The traditional concepts of agent environment and its constraints are expanded through the use of agent migration. Agents are capable of mobility between embodiments in virtual environments (e.g. virtual avatar), embodiments in physical environments (e.g. robot), and software environments (e.g. OS desktops, PDA's) (see figure 1). Once instantiated in the world, the agent has knowledge of that world, and of its capabilities therein.

Key technologies underpin such nomadic characteristics. These include white and yellow pages services for the location of people and services respectively. Agents would thus have access to directory services that permit access to publicly available resources and other agents. The agents would typically have a private directory for accessing resources specific to its owner and not publicly available. These agents act as a proxy for the user in the real and virtual worlds, as well as allowing the user remote access to their devices and public resources.

Figure 1. The Agent Chameleon spirit and its body instantiations: mobile devices (PDA), robot, PC, Web, and Virtual Reality respectively

Furthermore, the chosen embodiment of the agent must be capable of change, of agent mutation. This is particularly true in VR, where the agent is free of any constraints that exist in the real world. The agents must be capable of modifying their embodiment instantiation in response to the environmental and task specific events. For instance, in an outer space-like VR environment, the agent could adopt the persona of a rocket to allow it to fly, thereby facilitating human interpretation. The agent must exhibit the ability to dynamically select an appropriate form with associated functional portfolios. Additionally, the system must be extensible; it should be possible for the easy addition of new types of embodiment instantiations for different situations.

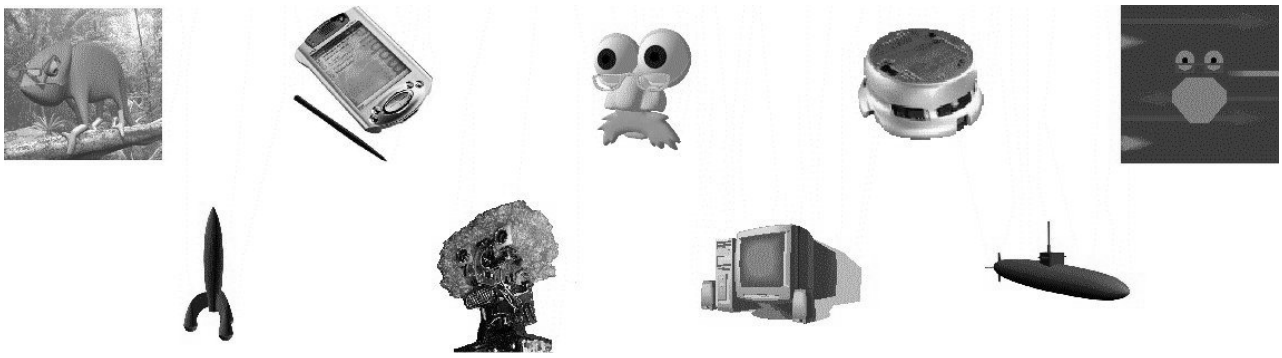


Figure 2. The current Agent Chameleon body instantiations as robots, VR avatars, entities on PC's and mobile devices

Alas, all platforms are not created equal (e.g. varying memory, processing power, bandwidth, display characteristics). Consequently, these agents have to be able to adapt to different conditions. Agents need to be able to evolve their very form. As they move from device to device they may necessarily have to shed some of their characteristics. This is analogous to exfoliation. Upon platforms that may not be able to handle an agent in its complete form, the agent is able to reduce elements that are non-essential to its task at hand and scale down its capabilities. Alternatively, on platforms with minimal resources, the agent merely sends a minimally sufficient component of itself, which is required only for its current task, while the bulk of the agent can remain dormant and dismembered on the source device. Upon task completion, both parts reintegrate and the agent continues on its way. Furthermore, an agent can, in certain circumstances, clone itself. Such circumstances may include those where it feels under threat, or under heavy resource demands. This elastic evolution would empower the agent with unforeseen versatility. The form of artificial evolution and adaptivity discussed is currently being developed within the Agent Chameleons framework (http://chameleon.ucd.ie). These concepts resonate with initiatives such as IBM's *autonomic computing* (Horn, 2001) and that of Intel's *proactive computing* (Tennenhouse, 2000)**.** Central to such initiatives is software comprised of confederations of autonomous and social agents which are capable of such facilities as self-healing, self-protection, self-configuration and self-optimisation.

With regard to embodiment issues, the body is always present and the reasoning of the agent is dependent upon that body as it provides the system with its actuator and preceptor functionality. However, the form of that body is not constrained; the agent is capable of adjusting it or adopting an entirely new one to suit the task at hand. Each different embodiment instantiation fundamentally changes the viewing metaphor for that agent and its associated functional portfolio. A key result of this work is that the issues of identity and association between the user and the agent chameleon are maintained through behavioural and visual cues as it migrates and mutates across platforms.

This work demonstrates how the traditional paradigm of one-mind-one-body can be extended beyond such a human-based reference to one-mind-many-bodies, thereby providing new core functionality for an artificial entity. Results to date have demonstrated the flexibility of an agent capable of migrating between platforms. Demonstrations at Media Lab Europe to visitors were found to successfully maintain the identity of the agent across platforms whilst employing the fundamentally different features of each as shown in figure 2 (physical mobility: Khepera robot, speech and facial gestures: Anthropos robot head; mutation and cloning: as facilitated through the virtual reality instantiations).

5. FORM OF THE MACHINE

The previous sections have discussed the changing function of the reasoning machine. With the permeation of computational devices in our society, the flexibility of artificial systems is changing. The next step is to look at how *we* will interact with these systems, how we will interact and understand these machines.

Conflicting arguments exist for and against the human form as a frame of reference for reasoning machines and these will continue to haunt robotics (see Duffy & Joue, 2004 for a discussion). The fundamental issue is how to achieve a balance between the function and form of the reasoning machine. Is the entity so strongly humanoid to the extent that we have the *replicant problem* as found in Philip K. Dick's famous novel: "Do Androids dream of electric sheep?" (1968). But, the functionality of the robot is then constrained by the human function and form. If aspects of the human form are used judiciously to facilitate human-robot social interaction (figure 3) (Duffy, 2003), its capability set can diversify from our own and embrace inherently mechanistic capabilities and possibilities (e.g. vision beyond human visible spectrum, wireless communication, multi-actuator derived degrees of freedom, auditory enhancement).

Figure 3. Media Lab Europe's "JoeRobot" at the Flutterfugue performance with SmartLab and NYU CATLab in London 2002 (Photo courtesy of Brent Jones)

The influence of the appearance and the voice/speech of an entity on people's judgements of another's intelligence have been demonstrated in experimentation. The more attractive a person, the more it facilitates others to rate the person as having higher intelligence (Alicke et al, 1986; Borkenau, 1993). However, when given the chance to hear the person speak, people appear to rate their intelligence more on verbal cues than on their attractiveness (Borkenau, 1993). Exploring the impact of such hypotheses to HCI, Kiesler and Goetz (2002) 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.

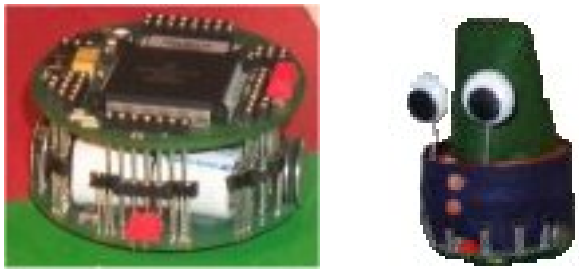
When we start to engage robots at a more complex level than our current interactions with washing machines, our propensity to anthropomorphise becomes inevitable. The important criterion is to seek a balance between people's expectations and the machines capabilities (Duffy, 2003).

The following case study explores our propensity to ascribe such notions as intelligence and emotions to machines.

5.1 CASE STUDY: Emotion Machines

The work presented in (Bourke & Duffy, 2003) demonstrates the ease with which people are willing to ascribe human-like characteristics such as emotion and intelligence to small robots performing computationally simple behaviours. This effectively highlights how much "mind" one is willing to ascribe to an artificial entity with little or no explicit design decisions involved.

The first stage of the experimentation involved the design and implementation of seven independent behaviours on standard Khepera I robots as shown in figure 4 (with some equipped with the wireless communication module). These were videotaped and a questionnaire was designed asking the observer to explain what the robots were doing, to pick three characteristics they would associate with the robots, and to grade these characteristics. Access to this questionnaire was distributed among a widely varied audience through Internet mailing lists. In the second stage, the same robots were dressed using coloured felt and given aesthetic "eyes" and the same behaviours were implemented. The same questionnaire was repeated.



(a) **(b)**

Figure 4. Experiments with Kepera I robots in (a) original form, and (b) in dressed form

The results indicated that in the first set of experiments, 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. However, it is useful to note that people seemed to easily see past the mechanics of the robots, and began to describe them as if they possessed some human-like qualities such as social interaction capabilities where no such explicit behaviours existed. The antenna on one of the robots was also interpreted as corresponding to a tail a number of times and some parallels were drawn with dog behaviour. In the second experiment, even stronger human-like features such as being 'alive' and 'playful' were reported. One interesting example, where one robot approaches an immobile second robot, moves around it, performs a shaking behaviour as if vying for attention, whereupon the other suddenly moves away, was explained in the context of the observer's interaction with their husband; on attempting to talk to him, he ignores her for a while, then just walks away.

This work raises the question of whether a system is required to be *inherently* intelligent or emotional in order for it to be *interpreted* as such. It is an orthogonal view of the pursuit of a system that one views as intelligent. An interesting aspect then arises. If the system can create the illusion of being intelligent and emotional, can it be maintained over time? Will its failing become apparent through our interaction with it? Similar to whether it *appears* intelligent or not, the issue of resolution will prevail. If the fake is good enough, we won't know the difference.

It is important to recognise Schneiderman's arguments against anthropomorphism (1988), which state that people employing anthropomorphism compromise in the design, leading to issues of unpredictability and vagueness. The argument effectively distils down to a distinction as to whether one can maintain the function of the robot as a tool or not. When actuation and perception mechanisms are employed on the system to engage with its physical and social environment, the notion of the system remaining purely a tool becomes less manageable and consequently anthropomorphism is unavoidable. It's how we manage the form and consequently the anthropomorphism that becomes the important issue (Duffy, 2003).

If we are so willing to ascribe standard social interaction frames of reference to clearly artificial systems, as demonstrated in these and other similar experiments, we should not fear developing technologies that clearly extend beyond our own capabilities as discussed in section 4. It is then the task of the designer to facilitate and maintain our "bond" with new future machines.

6. THE FUTURE MACHINE

It can be more tangible and manageable to use ourselves and similar standard paradigms as frames of references in designing artificial systems. But, as Einstein is reputed to have said, "*as far as the laws of mathematics refer to reality, they are not certain; as far as they are certain, they do not refer to reality*". There are many fundamental distinctions between artificial and natural systems. From time to time we need to stop, think laterally, try to free the artificial system and allow it exploit these differences. In the future, digital personal assistants will emerge to take advantage of their abstraction from a particular environment or platform and migrate, mutate, even clone as presented in the Agent Chameleon work discussed earlier.

Our perception of these systems may also change. Interestingly, when experiencing "Ada - l'espace intelligent" (Delbruck et al, 2003), a room where the user interacts with the room as much as the room interacts with the user, the aspect of the unknown and the elemental communication with this room helped created a strong sense of intelligence. Complexity is not necessarily the solution to creating an impression of a system being "intelligent" and this will influence the pursuit of a system that we will view as artificially intelligent.

A number of fundamental tenants that underpin this next generation of reasoning machine are considered in the following paragraphs. Key to these perspectives, as reinforced in the previous two case studies, are (a) the embracing of those features and capabilities inherent to artificial systems, and (b) the management of our willingness to anthropomorphise in our interactions with these systems.

6.1 Nomadic Agents

The Agent Chameleon work outlined previously, regarded the agent as an entity empowered with autonomy, human-computer interaction facilities, and a fundamental mobility. The embodiment instantiation thus merely becomes the container for a digital mind which opportunistically migrates between devices. The specifics of the hardware now reflect the capability set of the autonomous mind. The presence of the agent moving through cyberspace, as the user moves through physical space, allows the associated user to be available at anytime through the agent and vice versa. Such nomadic agents also have the capacity to exploit an *elastic cloning*. This involves the cloning of an agent into two or more agents for a particular task and then the "offspring" agents returning to the parent and all fusing together as one when the task is completed. Similar to platform migration, such temporary cloning is a concept that is facilitated through the technological advantages of software-based virtual systems, functionality with little to no basis in biology.

A clear application of such nomadic evolvable agents is that of an autonomous "intelligent" digital assistant that is independent of any one physical device. These entities will effectively give any user their own personal assistant that will help with the information overload in daily life, assisting with personal communications and offer a generic interface to any number of devices. They will have the ability to react to the current needs of there user, and beyond this, grow and learn to anticipate future needs and requirements. Perhaps our vision can be best summed up by Luc Steels' metaphor for what the robots of the future will be like: "*[it] is related to the age-old mythological concept of angels. Almost every culture has imagined persistent beings which help humans through their life. These beings are ascribed cognitive powers, often beyond those of humans, and are supposed to be able to perceive and act in the real world by materialising themselves in a bodily form at will."* He goes on to detail how Angels may *" project the idea of someone protecting you, preventing you from making bad decisions or actions, empowering you, and defending you in places of influence*" (Steels, 2000).

6.2 A Real Fake

When the fake is so good, we won't be able to tell the different between whether it is real or not. The Machiavellian Intelligence Hypothesis (see Kummer, 1997 for a recent discussion) proposes that intelligence as we understand it evolved from the social domain where social interaction between entities is key to the development of intelligence. It is because of our developing social interaction with machines, which are becoming more and more autonomous, that our perceptions of whether they are intelligent or not, or even *how* intelligent, becomes an issue.

Relative to our capability set, the idiosyncrasies of a robot embedded in our physical and social spaces equipped with such existing systems as flawed vision, annoying speech and woefully inadequate sensor systems could be the physical equivalent of a spam assault through chronic annoyances. They have their tragic flaws and may therefore become as alive as we are. This also raises an interesting point about machines as "constant companions"; what are the health and environmental drawbacks of having machines embedded in our physical and social space as autonomous entities? The solution to these problems is to define the task, the function, for the machine. This will dictate the form, which, in conjunction with the function, should embrace the fact that it is a machine, not confuse it.

7. CONCLUSION

The demands on machines like robots have dramatically increased during the last decades. No doubt, the film industry has contributed greatly in moulding the imagery with which we associate the reasoning machine. Often it is presented as a friendly, hard-working and droll creature, such as R2D2 in Star Wars (Lucas Films, 1977) and the more human-like character of Data in the Star Trek (1987) series. Creating an artificial being based on the blueprint of humans seems to be particularly compelling due, in part, to the basic effort of mankind to reproduce itself or even for the desire to be immortal. However, the film industry is not satisfied with presenting the bright side of machines. It also projects people's fears of machines into characters like The Terminator (MGM/UA, 1984), a nearly indestructible cyborg assassin.

This paper has sought to review and assess our perceptions of reasoning machines. Within this paper we have reflected upon the mind-body debate and the monistic versus dualistic standpoints. We have sought to extend the one-mind-one-body approach to accommodate a one-mind-many-bodies metaphor. Within this metaphor the agent's embodiment form may take many guises with the artificial mind or *agent* potentially exhibiting a nomadic existence opportunistically migrating between a myriad of instantiated embodiments.

The choice of embodiment must, not only empower the agent, but maintain the agent's identity in the eyes of the user. Central to this is the need to preserve key referential characteristics across the different instantiations. The title of this paper is intentionally *not* called "future *intelligent* machines". It is not the aim of the ideas proposed to argue against the necessity of embodiment in the pursuit of the artificially *intelligent* system, but rather to seek to take an orthogonal perspective and, whilst employing those technologies developed in the field of AI research, realise systems with new and fundamentally different capabilities. It is also questionable whether the term "intelligent" can be justifiably used in the majority of AI research to date with its interpretation being rather nebulous and vague.

We postulate a new generation of reasoning machines, which evolve and demonstrate autonomic characteristics. These machines are social (Duffy, 2000), autonomous, intentional and are equipped with rudimentary self-healing, self-protection, self-configuration and self-optimisation capabilities (note: these capabilities are better served in software rather than hardware – the use of software migration strategies unloads the hardware complexities required to achieve this functionality and invariably the chances of it failing in the first place). The scale of each of these features is primarily dependent on the complexity required and deployed, and draws on vast research to date which address these specific issues. The agent's sophistication is dependent on the extent of these technologies employed. The key feature, as presented in this work, is to highlight the fundamental perspectives that future reasoning machines can adopt.

While anthropomorphism in robotics raises issues about the taxonomic legitimacy of the classification *human*, and its sole association with ourselves, the question of whether machines will ever approach human capabilities persist. Technology is now providing robust solutions to the mechanistic problems that have constrained robot development thus far, thereby allowing robots to permeate all areas of society from work to leisure. The key is to take advantage of these reasoning machines and their capabilities rather than constrain them. We just keep in mind something like Asimov's Laws of Robotics (1994), and remember where the OFF button is.

ACKNOWLEDGEMENTS.

A sincere thank you to John Bourke who ran the experiments presented in (Bourke et al, 2003) and discussed in section 5.1). We also gratefully acknowledge the financial support of the Higher Education Authority (HEA) Ireland and the Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan.

REFERENCES

Alicke, M.D., Smith, R.H., & Klotz, M.L. (1986) "Judgments of physical attractiveness: The role of faces and bodies". Personality and Social Psychology Bulletin, 12(4), pp381-389.

Asimov, I. (1994) I, Robot, Bantam Books; Reprint edition (July)

Borkenau, P. (1993) "How accurate are judgments of intelligence by strangers?", Annual Meeting of the American Psychological Association, Toronto, Ontario, Canada, August

Bourke, J., Duffy, B.R., (2003) "Emotion Machines: Projective Intelligence and Emotion in Robotics", IEEE Systems, Man & Cybernetics Workshop (UK&ROI Chapter), Reading, September

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

Brooks, R.A., (1986) “A Robust Layered Control System for a Mobile Robot”, IEEE Jour. Rob. and Autom., 2(1)

Brooks, R.A., (1990) “Elephants Don't Play Chess”, Robotics and Autonomous Systems Vol. 6, pp. 3--15.

Brooks, R.A.,(1991) “Intelligence Without Representation”, Artificial Intelligence Journal (47), pp. 139-159

Delbruck, T and Eng, K and Baebler, A and Bernardet, U and Blanchard, M and Briska, A and Costa, M and Douglas, R and Hepp, K and Klein, D and Manzolli, J and Mintz, M and Roth, F and Rutishauser, U and Wassermann, K and Wittmann, A and Whatley, A M and Wyss, R and Verschure, P F M J Ada: a playful interactive space , Human-Computer Interaction --- INTERACT'03 : IFIP TC13 International Conference on Human-Computer Interaction, 1st-5th September 2003, Zürich, Switzerland. 989-992, Rauterberg, M et al. (Eds.), IOS Press, 2003

Dennett, D., (1987) The intentional stance, MIT Press, Cambridge, MA.

Dick, P.K. (1968) Do Androids dream of electric sheep?, Del Rey, Reissue edition (June 1996)

Descartes, R., (1637; Reprint Indianapolis: Cambridge Hackett Publishing, 3rd edition, 1993) Discourse on Method and Meditations on First Philosophy.

Dreyfus, H. (1972) “What Computers Can’t Do: The Limits of Artificial Intelligence”, MIT Press

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

Duffy, B.R., (2003) “Anthropomorphism and The Social Robot”, Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems 42 (3-4), 31 March, pp170-190

Duffy, B.R., Joue, G. “Embodied Mobile Robots”, 1st International Conference on Autonomous Minirobots for Research and Edutainment - AMiRE2001, Paderborn, Germany, October 22-25, 2001

Duffy, B.R., Joue, G., (2004) “I, Robot Being”, Intelligent Autonomous Systems Conference (IAS8) 10-13 March, The Grand Hotel, Amsterdam, The Netherlands

Duffy, B.R., O’Hare, G.M.P. , Martin, A.N., Bradley, J.F., Schön, B. (2003) “Agent Chameleons: Agent Minds and Bodies”, 16th International Conference on Computer Animation and Social Agents (CASA 2003), Rutgers University, New Jersey, May 7-9

Francher, R.E. (1979) Pioneers of Psychology, W. H. Norton & Company

Fukuda, T., et al. (1989) “Structure Decision for Self Organising Robots Based on Cell Structures”, IEEE: Rob. & Autom., Scottsdale Arizona

Horn, P., (2001) “Autonomic Computing: IBM’s perspective on the state of information technology”, IBM Corporation, Oct. 15th, http://www.research.ibm.com/autonomic/manifesto/autonomic_computing.pdf

Kiesler, S., Goetz, J., “Mental models and cooperation with robotic assistants”, Proceedings of CHI, 2002.

Kube, C.R., Zhang, H. (1993) “Collective Robotics: From Social Insects to Robots”, Adaptive Behavior, 2(2):189-219

Kummer, H., Daston, L., Gigerenzer, G., Silk, J., “The social intelligence hypothesis”, (1997) Weingart et al. (eds), Human by Nature: between biology and social sciences. Hillsdale, NJ: Lawrence Erlbaum Assoc., P157-179

Lucarini, G., Varoli, M., Cerutti, R., Sandini, G. (1993) “Cellular Robotics: Simulation and HW Implementation”, Proceedings of the 1993 IEEE International Conference on Robotics and Automation, Atlanta GA, May, pp III-846-852.

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

O’Hare, G.M.P. Duffy, B.R. Schoen, B., Martin, A.N. Bradley J.F. (2003) “Agent Chameleons: Virtual Agents Real Intelligence”, 4th International Working Conference on Intelligent Virtual Agents (IVA) 2003, 15-17 September, Kloster Irsee, Germany, LNCS Springer Verlag.

Searle, J. (1980) Minds, Brains, and programs, The Behavioral and Brain Sciences 3, 417-457.

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

Shneiderman, B. (1988) “A nonanthropomorphic style guide: Overcoming the humpty-dumpty syndrome”. The Computing Teacher, October, 9-10.

Steels, L. (2000) “Engeln mit Internetfluegeln. German version of Digital Angels”, In: Die Gegenwart der Zukunft , pp. 90-98, Verslag Klaus Wagenbach, Berlin

Sharkey, N., Zeimke, T., (2000) “Life, mind and robots: The ins and outs of embodied cognition”, Symbolic and Neural Net Hybrids, S. Wermter & R. Sun (eds), MIT Press

Thagard, P., (1996) “Mind, Introduction to Cognitive Science”, MIT Press

Tennenhouse, D.L., (2000) “Proactive Computing”, Communications of the ACM, 43, No. 5, 43-50, May

BIO's

Brian Duffy

Brian Duffy's research aims to understand man-machine interaction from the perspective of socially capable robots situated in the Media Lab Europe's office environment. This seeks to research the fine line between observed and designed function and form. It is an exploration of the *illusion* of life and intelligence in artificial entities.

Brian's interest in designing and building social humanoid robots has developed from previous research at the Department of Computer Science at University College Dublin where he completed a doctoral thesis. Prior to this, Brian spent two years in artificial intelligence research and building robot prototypes at GMD's (now the Fraunhofer-Gesellschaft Institute) Autonomous Intelligent Systems Institute. Before moving to Germany in 1994, two years were spent at Institut National des Sciences Appliquées de Lyon in France working in the field of Distributed Artificial Intelligence and Multi-Agent Systems.

Gregory O'Hare

Gregory O'Hare is Head of Department of Computer Science at University College Dublin (UCD). Prior to this he was a member of faculty at the University of Manchester Institute of Science and Technology (UMIST). He is director of the PRISM (Practice and Research in Intelligent Systems and Media) Laboratory within the Department of Computer Science. His research focuses upon Multi-Agent Systems (MAS) and Mobile and Ubiquitous Computing. He has published some 120 journal and conference papers in these areas together with two textbooks.

Gregory has secured some 6 million euro research funding for his research. He has acted as consultant for many national and international companies and organisations.

John Bradley

John Bradley received his Bachelor Degree in Computer Science in 2002 from the University College Dublin (UCD), Ireland. From there he joined the Agent Chameleons project (a joint collaboration between the Computer Science Department UCD and the Anthropos Group in Media Lab Europe) as a PhD Student.

John's current research involves the self-adaptation of agents, as they migrate between heterogeneous platforms, driven by deliberative mechanisms. His other research interests include agent technologies, multi-agent systems, distributed artificial intelligence, agent migration, agent adaptation and robotics.

Alan Martin

Alan Martin completed a Bachelor of Science Degree in Computer Science at University College Dublin, Ireland, and is now a PhD candidate there, working in collaboration with Media Lab Europe.

Alan's research interests include collaborative virtual environments, immersion, animated synthetic characters, agents and robotics.

Bianca Schoen

Bianca Schön completed a Bachelor of Science Degree at the University of Applied Sciences in Darmstadt, Germany.

During her studies she gained experiences not only in economic enterprises like T-Systems debis Systemhaus GmbH in Darmstadt and the H.A.S.E GmbH in Hünfelden, but also in scientific institutes like the Frauenhofer Institut für Grafische Datenverarbeitung and the T-Systems Nova Technology Center, both situated in Darmstadt, Germany.

She is now a Ph.D. candidate at the University College Dublin, Ireland, where she is working in a collaboration project with Media Lab Europe.

Bianca's research interests include animated synthetic characters, artificial intelligence, agent technology, evolutionary- and genetic programming and evaluation methodologies.