Towards Social Intelligence in Autonomous Robotics: A Review

BRIAN R. DUFFY
Media Lab Europe
Sugar House Lane, Bellevue, Dublin 8,
IRELAND
brd@media.mit.edu http://www.medialabeurope.org/people/b-duffy.html

Abstract: - This paper aims to provide grounding for the development of social intelligence in autonomous mobile robots by highlighting that little has been done in developing an explicit social framework for a team of robots engaged in collaborative work. Only recently has work been undertaken on developing social attributes and functionality in robotics [23] [16]. This review provides a somewhat historical perspective of artificial intelligence research in autonomous mobile robotics from classical approaches through "New AI" to a discussion of the foundation issues for implementing a degree of explicit social functionality.

Key-Words: - Social robots, social intelligence, embodiment, emotion

1 Introduction

In order to understand the issues involved in designing and implementing a methodology for the social interaction of a team of robots in real-world scenarios, and as such take multi-robot research a step further, an number of fundamental issues arise. This paper starts with a somewhat historical perspective to the development of artificial intelligence from the classical approaches to current ideas and trends regarding adaptive systems. The notion of "embodiment" is then discussed in detail regarding current interpretations and misinterpretations in the robotics domain. The embodiment of a robot is extended from its traditional physical interpretation to the social domain for real-world robots. This provides a basis for the development of an understanding of intelligence based on the implications of sociality and social "situatedness" in artificial intelligence research. Finally, the attributes associated with sociality including language, communication, and emotion are briefly discussed.

2 Classical to New AI

Over the last decade there has been a significant development in the Artificial Intelligence research community regarding the concept of embodiment and the development of an artificially intelligent robot. Two distinct methodologies have developed regarding *representation* vs. *perception*. Traditionally, artificial intelligence has developed from the representational perspective based on world modelling with a change in the late 1980's to work on perception in an attempt to combat failings of representational methods. The following sections briefly outline the evolution of AI research from classical symbol manipulator algorithms to embodied Artificial Life experiments with autonomous mobile robots.

2.1 "Good Old Fashioned AI"

The thesis of *Good Old Fashioned Artificial Intelligence* (GOFAI) is that the processes underlying intelligence are symbolic in nature. A Turing model [67] [68] of intelligent behaviour, viewed as essentially computational, inspired these first steps into the development of "artificial intelligence". More specifically, GOFAI models human intelligence as von Neumann computational architectures that perform computations on abstract symbolic representations. These computations are governed by a stored program, which contains an explicit list of instructions or rules. These rules transform the symbolic representations into new symbolic states. As such, GOFAI depicts mentality within the context of what philosophers know as the *Representational Theory of Mind* (see [60] for a recent discussion), according to which the mind is an entity that performs calculations on mental representations or symbols, which refer to features of the outer world:

> *"Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures"* Paul Thagard [64]

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.

While this hypothesis is strictly scientific, it does not take the current expansion of the cognitive science umbrella to encompass embodiment, albeit either physical or social. The principle drawback of the classical approaches to artificial intelligence as a control paradigm for robots is that explicit reasoning about the effects of low-level actions is too computationally expensive and time consuming to generate real-time behaviour. Another is the perception complexity problem, where noise and errors in the perceived environment result in decisions based on incorrect perceptions. As the environment itself increases in complexity, its correct perception becomes even more difficult.

2.2 The Real World and New AI

The fundamental difference between the *representation* (or "Classic AI") and *perception* (or "New AI") based approaches lies in the degree of interaction between the "body" and the "brain". While two communities battle over the validity of seeking more exact world representations, intuition leads many of us to the conclusion that approaching the problem of robot control from the perspective of strictly structuring, exacting, and continuously quantifying with increasing accuracy, is not the only possibility. This only provides a digital solution to an analogue problem that suffers from ever increasing complexity. Physical embodiment *necessitates* the use of approximate solutions as such solutions are inherently based on noisy and often incorrect perceptions. As yet, the relationship between body and intelligence continues to create a lot of debate and food for thought with few claiming that there is no relationship between the two [10] [3] [61] [32] [20]. Embodiment constitutes system-environment interaction and is dealt with in greater detail in section 3.

Other than the aspect of exact solutions being very computationally demanding, they also may not even be valid to the reality of the application as the solution is quickly outdated in dynamic environments. A simple example of this would be a mobile robot undertaking a docking procedure for recharging. Odometric errors and the robots inability to drive exact and correct paths demonstrate where an exact path plan is both unnecessary and unachievable. This exemplifies the differences (and problems) that exist between *representation* and *perception* based methodologies.

The term *New Artificial Intelligence* or *Nouvelle AI* is a recently coined term and has been used by researchers like Pfeifer *et al.* [47] [45] in discussing embodied cognitive systems and in particular mobile robots. *New AI* is a new methodology for studying intelligence and for understanding the mind with a view to providing a framework for alternative approaches to the classical stance. One of the main characteristics of *New AI* is its investigation of system-environment interaction. Although neuroscience, and in particular the field of neural information processing, has a bias towards information processing, it is becoming ever more obvious that there are two dynamics, namely the control architecture, e.g. neural networks, and the physical system. When integrated properly, there can be cooperation between the two, which results in control architectures utilising certain environmental properties to their benefit. Such approaches have lead to emergent robot behaviours where often quite simplistic control architectures can display relatively complex behaviours [41] [48] [37]. Researchers working within this paradigm have not hesitated to term some such emergent behaviours as "intelligent". While it can be extremely difficult to reproduce such behaviours explicitly, the paradigm does not facilitate the development of explicit complex behaviours. While interesting, inherent problems regarding recordability, observability, repeatability, and analysis (both quantitative and qualitative) challenge this approach from a scientific perspective. In contrast, research on *top-down* control via deliberative architectures has displayed high-level reasoning capabilities but lacks real world robustness.

The inherent problems with existing robot control approaches therefore reinforces the importance of understanding embodiment.

3 Physical Embodiment

René Descartes is referred to as the father of cybernetics due to his study of the human body as a machine. Descartes, in *Meditations* [17], aimed to show that mind is distinct from body. He points out that even though he may have a body, his true identity is that of a thinking entity alone and, indeed, his mind could exist without his body. He argued 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. Sense perception, movement, and appetite may require a body but they are only attributes of our body and not of our spirit and, hence, do not comprise our essence.

While some treat the body as peripheral and tangential to intelligence, others argue that

embodiment and intelligence are inextricably linked [3] [61]. In contrast to the representational CRUM perspective, Brooks popularised the claim by the German philosopher Heidegger [21] that we function in the world simply by being part of it. Brooks uses the phrase "being-in-the-world" in terms of his implementation of the subsumption architecture to autonomous mobile robots. Experience in building robots has led Brooks to argue that embodiment is vital to the development of artificial intelligence [5] [6]. Brooks advocates the behaviourist approach to combat the difficulty in developing purely internal symbolic representational models of reality utilised in classical AI approaches.

Lakoff *et al.* argue that our ability to understand and reason abstractly relies heavily on our bodily experience and that "high level" intelligence depends crucially on embodiment [35] [32]. Based on the argument of movement, manipulation and perception involving the use of recurring patterns, this promotes the concept of linking embodiment to intelligence. Phenomenologists also argue against the use of internal symbolic representations or mental states saying that *"an embodied agent can dwell in the world in such a way as to avoid the…task of formalising everything"* because its *"body enables [it] to by-pass this formal analysis"* [20]. Dreyfus also says that when people have "*mental considerations*", they *"do so against a background of involved activity"* [21].

Clark uses the term *blueprints*, indicating a highly detailed plan or specification, while discussing cognition and specifically "embodied cognition" in relation to the developmental process in infants, according to which "*mind, body and world act as equal partners*" [10]. Clark follows the notion that embodiment is crucial to intelligent systems, which research has traditionally tended to dissect.

Embodied cognition is unique for all natural systems. This is due to the individual experiences collected during a system's lifetime. It is little argued that intelligent systems are required to have some learning from experience mechanisms in order to function in complex non-deterministic environments. The system must be able to update and add to its knowledge set in order to survive. The "Artificial Life" or "Alife" community has approached the notion of a robot "surviving" from an alternative perspective (see [56] for an introduction). Artificial Life involves the embodiment of robots in an environment with the principle of surviving for a period of time, generally a time scale measured as a multiple of the robot's battery life (i.e. 30x battery life). Artificial life has been defined by Langton as:

> *"The study of man-made systems that exhibit behavio[u]rs characteristic of natural living systems. It complements the traditional biological sciences concerned with the analysis of living organisms by attempting to synthesi[s]e life-like behavio[u]rs within computer and other artificial media."* C. G Langton [33]

Olson [44] discusses the notions of "weak artificial life" and "strong artificial life" by differentiating *weak alife* as being the use of computers to simulate life, and *strong alife* as the claim that *"computer programmers can, at least in principle, go beyond mere modelling and literally create living things*". Olson does not discuss the use of physical robots, but rather seeks to argue that computer-generated organisms are material objects. This is discussed from a very philosophical perspective and lacks foundation in real world concepts and applications. While some research has been conducted in simulators and purely software-based systems, the real challenge lies in physically embodied artificially "alive" entities.

The following sections seek to address the inherent problem highlighted in [62] where "*sometimes it is unclear in the literature whether it is the controller that is embodied in the robot or the robot that is embodied is the world*".

3.1 Strong Physical Embodiment

While some believe that implementing a control paradigm on a physical robot is sufficient for fulfilling the embodiment criteria, Dautenhahn and Christaller [15] argue that this results in a robot not being aware of whether it is acting in a simulated or physical body. They write that the "*development of a* conception *of the body, which is generally discussed as the acquisition of a* body image *or* body schema, *is necessary for embodied action and cognition*". They continue in proposing that the use of evolvable robots with an adaptation of both body and control mechanisms to its environment could provide an ideal solution.

Maturana and Varela [40] differentiate between this issue of animal systems versus mechanical systems by concentrating on the organisation of matter in systems (see also [63]) via the terms *autopoiesis* and *allopoiesis*. In essence this constitutes the fundamental distinction between true embodiment and an artificial intelligence perspective of embodiment. *Autopoiesis* means self- (auto) creating, making, or producing (poiesis). Animal

systems adapt to their environment at both macro (behavioural) and micro (cellular) levels and are therefore termed autopoietic systems. Mechanical systems on the other hand can only adapt at a behavioural level and are termed allopoietic.

Similarly, Sharkey and Zeimke highlight in [63], "[*l*]*iving systems are not the same as machines made by humans as some of the mechanistic theories would suggest*". The fundamental difference lies in terms of the organisation of the components. Autopoietic systems are capable of self-reproduction. The components of a natural system can grow and ultimately grows from a single cell, or the mating of two cells. In such systems, the processes of component production specify the machine as a unity.

Allopoietic systems are, on the other hand, a concatenation of processes. Its constituent parts are produced independently of the organisation of the machine. This fundamental difference, in the context of artificial intelligence, has been highlighted in [62] where the notion of evolvable hardware is discussed. The designer of a robot is constrained by such issues as the physical and chemical properties of the materials used, by the limitations of existing design techniques and methodologies. The introduction of evolvable hardware could help overcome the inherent global limitations of the robot end product by facilitating adaptation and learning capabilities at a hardware level rather than only at a software level. This adaptability is often taken for granted in biological systems and likewise ignored when dealing with such issues as robustness, survivability, and fault tolerance in robotic systems. Sharkey and Zeimke highlight the lack of evolvable capabilities in allopoietic systems as being directly related to its autonomy, i.e. it is not. Biological or autopoietic systems *are* fully autonomous.

While embodiment has been approached from different perspectives by the mentioned authors, the conclusion is similar. Embodiment is an inherent property of an agent that exhibits intelligent behaviour leading to the now established hypothesis that, in order to achieve cognitive capabilities or a degree of intelligence in an agent, a notion of embodiment is required where there is interaction between "body" and "mind".

4 Social Embodiment

To date, a fundamental facet of embodiment has been neglected in autonomous mobile research and in artificial intelligence as a whole, that of *social embodiment*. Embodiment has been interpreted as being the physical existence of an entity, i.e. a robot, in a physical environment (by robot it is understood to represent a physical body with actuator and perceptor functionality). By virtue of it physical presence, whether static or dynamic, there is interaction between the robot and the environment. At a fundamental level, this can be the physical space occupied by the robot and extending to the robot's ability to move, change, and perceive the environment. When a second robot is added, this introduces a definite element of social interaction, even without any direct inter-robot communication. The perceptions of another robots motions, whether abstract notions of a moving obstacle or its clear distinction as another individual robot, influences the observing robots behaviour. The social implications of two robots coexisting in an environment add another dimension to the complexity of each robot's environment, which cannot be ignored.

In the first instance where the first robot perceives a moving obstacle in some simplistic way, the overlap between the concept of embodiment and the "social" connotations of one robot's influence over another becomes apparent. While this abstract notion of "communication" is not explicit in either its intention or application, it does not constitute a "degree of sociality" in its strictest sense.

4.1 The Machiavellian Intelligence Hypothesis

According to the *Machiavellian* (or *Social) Intelligence Hypothesis*, primate intelligence originally evolved to solve social problems and was only later extended to problems outside the social domain [11] (recent discussion in [30]). An example is where monkeys, which can exhibit a capacity to deal with complex social problems, are unable to transfer and adapt knowledge from one domain to another. Humans obviously can and are inherently linked to abstraction, generalisation and analogical reasoning capabilities.

Humans are also aware of their mental states (i.e. motives, beliefs, desires, and intentions) and can attribute mental states to others, which allows one to predict and analyse the behaviours of both oneself and others. This allows one to be able to deal with both highly complex social relationships and also exhibit the ability to deal with abstract problem solving. The social intelligence hypothesis claims that all these intellectual capacities evolved out of a social domain, i.e. out of interactions with socially embodied individuals.

This promotes the theory that in order to achieve a degree of intelligent behaviour from an agent, the agent must be both embodied in a physical environment and embodied in a social environment. This agent would therefore be subjected to complex dynamic social interactions in a real world, which are also believed by [15] to be necessary for the development of an artificially intelligent agent.

In addressing the issue of social embodiment, the notion of social intelligence must be discussed.

5 Social Intelligence

Psychologists have been defining other intelligences for some time, and grouping them mainly into three clusters: *abstract intelligence*, *concrete intelligence*, and *social intelligence* [52]:

> *Abstract intelligence*: the ability to understand and manipulate verbal and mathematic symbols.
> *Concrete intelligence:* the ability to understand and manipulate objects.
> *Social intelligence:*the ability to understand and relate to people.

Howard Gardner [25] proposes a theory of *multiple intelligences*, based on biological as well as sociological research, and formulates a list of seven intelligences. These include: *logical-mathematical intelligence*, *linguistic intelligence*, *spatial intelligence*, *musical intelligence*, *bodily-kinaesthetic intelligence* and *personal intelligence*. This last category includes two separate intelligences: *interpersonal intelligence* (the ability to understand the feelings and intention of others and *intrapersonal intelligence* (the ability to understand one's own feelings and motivations).

E.L. Thorndike identified the concept of "social intelligence" in 1920 [65] and defined social intelligence as:

> *"the ability to understand and manage men and women, boys and girls -- to act wisely in human relations."* E.L. Thorndike [65]

Watson *et al.* [69] include inter- and intra-personal intelligences in his theory of multiple intelligences. These two aspects comprise social intelligence. They define them as follows:

> ***Interpersonal*** *intelligence is the ability to understand other people: what motivates them, how they work, how to work cooperatively with them. Successful salespeople, politicians, teachers, clinicians, and religious leaders are all likely to be individuals with high degrees of interpersonal intelligence.*
> ***Intrapersonal*** *intelligence ... is a correlative ability, turned inward. It is a capacity to form an accurate, veridical model of oneself and to be able to use that model to operate effectively in life.* Watson et al. [69]

Social intelligence entails the skills and abilities involved with creating and maintaining a community. Jane Braaten points out that in many West African languages our word intelligence translates into a word that "denotes social skills and in particular, social abilities that are strongly associated with power to contribute to society" [2]. The building of a community she says "requires special intellectual competence". This competence is social intelligence and can be defined as:

> ***Social Intelligence:*** *The intelligence that lies behind group interactions and behaviours.*

Worden in [72] proposes a computational theory of primate social intelligence, in which primates represent social situations internally by discrete symbol structures, called scripts. This theory is compared with primate data from an intuitive perspective with possible experiments discussed. Such experiments are based on constructing the sum social knowledge for a primate species, and its expression in scripts. Worden makes some fundamental assumptions regarding the other components of the brain. The first is that "*there is an internal representation, or mental model, of Local Space and Motion*". The second assumes that "*feature individuation and categorisation are solved by other modules in the brain*", which deliver categorised, individuated symbols to the social intelligence module (SIM). While such a social cognitive model may be useful for studying social interaction in primates, it adopts a classical AI stance regarding internal environment models. Worden notes that "*[t]his assumption is doubtless an approximation, but is a necessary one in order to proceed to a first understanding of the SIM*".

Worden [72] discusses the structure of the social domain as consisting of the following:

- The structure and interrelations between the components are crucial.
- The set of social situations and possible causal relations between situations are systematic sets.
- The set of possible situations is very large
- An agent's social milieu involves discrete, identified individuals who tend to be in discrete relations to one another.
- The interval between social cause and effect may have extended time frames.
- Generalisations across individuals are important (i.e. standard social responses and interpretations)

- There is a chaining of cause and effect: if A causes B, and B causes C, then effectively, A causes C.

Formal computational descriptions of primate social intelligence have been proposed by Byrne [9], Schultz [54] and Schmidt and Marsella [59]. These are mainly concerned with the high order problems of recognising agency and another agent's plans within a primate theory of mind. Byrne addresses the issue of the first-order problem of primate social intelligence (without a theory of mind) via a production rule formalism but has not been extended (yet) to include a worked out theory of learning tailored to the social domain. In exception to learning, there are considerable similarities between Worden's scripts and Byrne's production rule formalisms.

5.1 The "Social Level"

Newell's [43] unified theory of cognition identifies a separate level above the rational level (the human equivalent of the Knowledge Level) for dealing with social contexts. He terms this the *Social Band* and serves to define an individual's behaviour in a social context. Newell acknowledges that his *Social Band* is not clearly defined, but that it should only contain knowledge that is specifically social in nature. Newell clearly states that "*there is no way for a social group to assemble all the information relevant to a given goal, much less integrate it*" and that "*there is no way for a social group to act as a single body of knowledge*".

Jennings *et al.* introduce the *Social Level Hypothesis* [28] to provide an abstract categorisation of those aspects of multi-agent system behaviour that are inherently social in nature, i.e. co-operation, co-ordination, conflicts, and competition. The *Social Level* sits above Newell's *Knowledge Level* (KL) [42]. The *Social Level Hypothesis* states:

> *"There exists a computer level immediately above the KL, called the Social Level, which is concerned with the inherently social aspects of multiple agent systems"* Alan Newell [42]

Jennings *et al.* discusses the *Social Level* from the context of social responsibilities and leads to the formulation of the *Principle of Social Rationality*.

5.2 Principle of Rationality and Social Rationality

It is believed that by explicitly drawing out a few key concepts that underpin the behaviour of a large class of problem solving agents, it is possible to understand and predict agent behaviour. Newell [42] proposed that the agent problem solving behaviour could be characterised through the *Principle of Rationality*:

> *"If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action"*

Jennings *et al.* [28] discuss the implications of this within a social context and have ascertained that for a number of social actions where there is a conflict of interest between that of the member and of the society itself, Newell's *Principle of Rationality* is flawed. Jennings *et al.* propose the *Principle of Social Rationality* as:

> *"If a member of a responsible society can perform an action whose joint benefit is greater than its joint loss, then it may select that action"*

The justifications for the extension of Newell's original proposal to the *Principle of Social Rationality* is based on the balance between the individual benefit between a members interests and those of the society or vice versa.

Jennings *et al.* continue by defining the minimum set of necessary concepts for a responsible society to obtain the behaviour specified by the *Principle of Social Rationality*:

- Acquaintance: the notion that the society contains other members
- Influence: the notion that members can effect one another
- Rights and duties: the notions that a member can expect certain things from others in the society and that certain things are expected of it by the society

Based on these fundamental attributes, a social community of agents can therefore exist. In order to achieve these attributes, the "tools" facilitating social interaction such as communication are required.

6 Social Interaction

A mutual understanding between members of a social group is required for the establishment and continuance of the group. Problems arise when single members pursue their own interests. The question of individual safety is only ensured when a strict and fixed hierarchy exists like in a dictatorship society or "super-organism" structure (social insects) with anonymous contact between its group members. But in most social groups degrees of flexibility are found with each member's social status either updated or confirmed regularly. While

non-human primate societies generally use physical contact for "social grooming" with the use of its body and behaviour for communication, man has developed a more effective means via a highly elaborate language.

Most higher vertebrate species learn to adapt their behaviour not in total isolation but in social groups. Social learning not only refers to general learning processes that are employed for group behaviour, but also refers to new and unique strategies to control the relationships between conspecifics. Searle [55] discusses the ability of minds, acting in cooperation, to create an objective social reality.

6.1 Communication and Language

Most living societies communicate, from insects to humans. In low-level species, communication abilities often appear built-in. However, in more complex creatures this capacity develops during their life. Studying systems that would develop such ability is a necessary and natural step towards the construction of robots demonstrating more complex cognitive capacities.

Existing communication methodologies can often be classified as either using explicit communication to pass state information between component robots, or, in the context of a reactive robot's control structure, build rules for such interactions.

Explicit communication can unfortunately fail when confronted with no predefined communication protocol or when external dissimilar agents are introduced into the system. Reactive approaches on the other hand, while robust and flexible in their limited domains, are too simplistic for use in domains that require more complex reasoning. It can be difficult to extend reactive systems to complex domains, and they frequently suffer from dysfunctional emergent behaviour when their rule-bases become complicated.

It is believed that language not only functions to acquire knowledge about behavioural characteristics of others, but also to find out their internal states (i.e. their feelings, attitudes, etc). In order to build up a basis for interaction and co-operation, individuals have to communicate and merge their conceptions of the world (i.e. world models), with the degrees of abstraction being socially grounded and continuously updated.

In addition to being perceived as a means of communicating ideas, knowledge and experience, language is also a tool where it alters the nature of the decision making process [38].

Communication in multi-robot systems ranges from none at all to interaction through high-level agent communication languages (ACLs). Among the approaches for robot communication are:

- None [31]
- Implicit communication through the environment [18] [19].
- Simple semaphores or signals. For example, in a reactive agent system a certain stimulus may trigger an agent to emit an alarm [1].
- Simple message passing [1].
- Agent Communication Languages [34] [24] [10] [51].

The level of communication complexity has generally been representative of the complexity of the robot systems in terms of both its interactive and computational capabilities. As few have negotiated the task of developing socially capable robot entities, work to date on communication has been primarily from a multi-agent systems perspective.

6.2 Degrees of Sociality

Agents existing in a social environment may have many different degrees of social interaction. The following gives examples of how different socially situated agents may cooperate:

Benevolent agents: accept all goals, which they are capable of performing, asking in the interest of helping… basically having the interests of the other robot in mind, being sympathetic, the assumption that agents do not have conflicting goals, and that every agent will therefore always try to do what is asked of it [49].

Altruist (selfless) agents: only perform goals, which contribute to the net benefit to the society, the greater emphasis is on its social utility.

Contrary to research on the competitive nature and interaction of robots, altruist co-operation between robots is based on one robot's unselfishness towards the interests of another.

Socially responsible agents: strikes a balance between altruism and selfishness.

Independent (selfish) agents: Stand-alone self-interest agents (solipsism: self knows nothing but its own modifications and the self is the only existent), more emphasis on individual utility.

Antagonistic agents: Forcing others to do what you want… territorial, i.e. forcing another robot out of the way at a recharging station.

Empathy: understanding and entering into another's feelings, or as a way of perceiving and comprehending the possible negative and positive experiences of another in a social situation involving

two agents [70] [7].

The term sympathy, while closely related to empathy, refers to the awareness and participation in the suffering of another, while empathy refers to the attempt to comprehend either positive or negative states of another. Wispe describes the difference this way:

> *"In empathy the self is the vehicle for understanding, and it never loses its identity. Sympathy, on the other hand, is concerned with communion rather than accuracy, and self-awareness is reduced rather than enhanced....In empathy one substitutes oneself for the other person; in sympathy one substitutes others for oneself. To know what something would be like for the other person is empathy. To know what it would be like to be that person is sympathy. In empathy one acts "as if" one were the other person.... The object of empathy is understanding. The object of sympathy is the other person's well-being. In sum, empathy is a way of knowing; sympathy is a way of relating" Lauren Wispe [71]*

Sympathy is performed with altruistic ends, but empathy may or may not be motivated by good intentions. Brothers [4] considers empathy as a form of "emotional communication" or the "reading" of social signals in humans and as being primarily a biological phenomenon. In [14] Dautenhahn discusses the concept of empathy and speculates that "*experimental, empathic understanding might be a mechanism which has not developed 'normally' in autistic people*" from the perspective of their social defects and that this "*could answer the question whether universal mechanisms of social behaviour exist*". The (re)search continues.

As the current state-of-the-art in socially developed robot systems is very limited, these degrees of social interaction has only recently been developed from an artificial intelligence perspective with a view to its application in the robotic domain [23]. This requires the development of a complete system of control enabling such degrees of social interaction, from a control architecture, language, and hardware point of view, which this work aims to address.

Such social attributes as empathy, sympathy, antagonism and etc are inherently associated with the notion of emotion.

7 Emotional Intelligence

Emotional Intelligence (EI) has its roots in the concept of social intelligence. Salovey and Mayer published the first formal definition of emotional intelligence in 1990:

> *"a type of social intelligence that involves the ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions"* Peter Salovey and John Mayer [58]

According to Salovey & Mayer [39], EI subsumes Watson et al.'s [69] inter- and intra-personal intelligences, and involves abilities that may be categorized into five domains:

- *Self-awareness:* Observing yourself and recognizing an emotion as it happens.
- *Managing emotions:* Handling emotions so that they are appropriate; realising what is behind an emotion; finding ways to handle fears and anxieties, anger, and sadness.
- *Motivating oneself:* Channelling emotions in the service of a goal; emotional self-control; delaying gratification and stifling impulses.
- *Empathy:* Sensitivity to others' emotions and concerns and taking their perspective; appreciating the differences in how people feel about things.
- *Handling relationships:* Managing emotions in others; social competence and social skills.

Mayer and Salovey [39] have recently updated their own definition of EI by stating that their previous (and other) definitions only dealt with perceiving and regulating emotion, and omit thinking about feelings.

> *"Emotional intelligence involves the ability to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge; and the ability to regulate emotions to promote emotional and intellectual growth".*Mayer and Salovey [39]

Neurologist and philosopher Antonio Damasio [12] draws an intimate connection between emotion and cognition in practical decision-making. Damasio presents a "somatic marker" hypothesis that explains how emotions are biologically indispensable to decisions. His research on patients with frontal lobe damage indicates that feelings normally accompany response options and operate as a biasing device to dictate choice.

Darwin highlighted the role the emotions play in the decision process when he wrote:

> *"I put my face close to the thick glass plate in front of a puff adder in the Zoological Gardens, with the firm determination of not starting back if the snake struck at me; but, as soon as the*

blow was struck, my resolution went for nothing, and I jumped a yard or two backwards with astonishing rapidity. My will and reason were powerless against the imagination of a danger which had never been experienced." Charles Darwin, [13]

Damasio offers convincing examples and arguments for a very deliberate attention to the body, in particular emotions, when reflecting on reasoning and decision-making. He explains, "*that our most refined thoughts and best actions, our greatest joys and deepest sorrows, use the body as a yardstick*". We live our lives, and experience our environments with our body as an active participant. The body is the vehicle through which we live all sorts of experiences.

"... love and hate and anguish, the qualities of kindness and cruelty, the planned solution of a scientific problem or the creation of a new artefact are all based on neural events within a brain, provided that brain has been and now is interacting with its body." Antonio Damasio [12]

Damasio came to this discussion through scientific explorations and findings. Damasio's work as a neurologist with patients suffering a myriad of disorders, including brain damage, and or problems with memory, language, and reason have led him to believe that mental activity requires participation from both the brain and the body. There are two relevant cases, that of Phineas Gage and "Elliot". Gage and Elliot were not able to relate socially, and yet all standard measurements of what is considered "intelligent" indicated that they were normal. Damasio, in effect, argues that the capacity to be emotional is synonymous with being socially intelligent.

DECO [8] is a computer program, which provides a means of testing out whether the principles of deliberative coherence can fruitfully be applied to understand real cases of complex decision-making. Allison Barnes and Paul Thagard proposed that the combination of Damasio's "somatic markers" with DECO provide the following sketch of a possible theory of emotional decisions:

1. Decisions arise when new information is inconsistent with one or more currently held goals. The mismatch yields a negative emotion, which produces a rupture in ordinary activity.
2. The decision juncture causes a simulation to occur, in which goals are re-evaluated on the basis of new information. This evaluation of goals elicits somatic markers.
3. Once somatic markers prioritise the goals, new options are simulated and evaluated.
4. Coherence calculations produce the best option and equilibrium is restored between the present situation and existing goals.

Based on these assumptions Barnes and Thagard propose that *"emotions function to reduce and limit our reasoning, and thereby make reasoning possible"*. It is believed that emotions prioritise thinking by directing attention to important information and that emotional states differentially encourage specific problem-solving approaches such as when happiness facilitates inductive reasoning and creativity. For a description of DECO and a comparison with classical decision theory, see [66].

Sloman [57] posed the question "could a goldfish long for its mother?" and argued that this would require rich cognitive structures and processes:

- Have a representation of its mother
- Believe that she is not in the vicinity
- Be able to represent the possibility of being close to her
- Should intrude into and interfere with other activities

This example demonstrates a purely internal driving force for a particular emotion, which Damasio [12] highlights in his distinction between primary emotions that are triggered by external or internal stimulation of various sense organs, and secondary emotions that are triggered by purely cognitive events.

In seeking an explanation of what an emotional state is, Sloman and Croucher [53] have proposed the following:

"a desire for something to be the case or not be the case: past present or future. It may be currently active, or dormant" Sloman and Croucher [53]

Examples of such being: anger, fear, delight, pity, awe, embarrassment, shame, pride, etc. Wright *et al.* [73] implements a self monitoring "meta-management layer" above a control architecture similar to Georgeff's Procedural Reasoning System [26] [50] which supports having and losing control of thoughts and attention via emotional stimulus: i.e. feeling ashamed of oneself or humiliated. Examples include aspects of grief, anger, excited anticipation, pride, and many more humanlike emotions.

Daniel Goleman, a clinical psychologist, defines EI as:

"the ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions" Daniel Goleman, [27]

Goleman discusses five basic emotional competencies: self-awareness, managing emotions, motivation, empathy and social skills in the context of emotional intelligence. This approach is a departure from the traditional attitude, still prevalent, that intelligence can be divided into the verbal and non-verbal (performance) types, which are, in fact, the abilities that the traditional IQ tests assess. Emotional intelligence, as defined by Goleman can be interpreted as being the personal intelligences observed by Gardner [25] (*section 5*).

7.1 Emotions and Cognitive Systems

There are numerous other theories, with new ones being refined regularly. Current thinking is that emotion involves a dynamic state that consists of both cognitive and physical events. Conversely, it can be argued that our everyday attributions of emotions, moods, attitudes, desires, and other affective states implicitly presuppose that people are information processors.

One could pose the question whether the attribution of artificial emotions to a robot analogous to the Clever Hans Error [46] where the meaning and in fact the result is primarily dependent on the observer and not the initiator? Possibly not. As emotions are fundamentally social attributes, they constitute communicative acts from one person to another. The emotions have a contextual basis for interpretation, otherwise the communication act of expressing an emotion may not be successful and may be misinterpreted.

As the case of Phineas Gage in 1848 demonstrated, an emotional capacity is fundamental towards the exhibition of social functionality in humans. Salovey and Mayer have reinforced the conclusions of Damasio that emotionality is an inherent property of social intelligence in discussing emotional intelligence. When coupled with Barnes and Thagard's hypothesis that *"emotions function to reduce and limit our reasoning, and thereby make reasoning possible"*, a foundation for the implementation of an emotional model in the development of a social robot becomes apparent. While numerous often-contradictory theories of emotion exist, it is accepted that emotion involves a dynamic state that consists of cognitive, physical and social events.

8 The Question of Being Social

Developing from Thorndike's original 1920's definition of social intelligence, through Watson's use of inter- and intra-personal intelligences to today's more computationally tractable interpretation, social intelligence and its preliminary application to robot communities is becoming feasible [22].

The *Cognition and Effect* work by Sloman *et al.* [36] aim at understanding the types of architectures that are capable of accounting for the whole range of human mental states and processes, including not only intelligent behaviour but also moods, emotions, desires, and etc. They propose that human-like architectures require several different sorts of concurrently acting sub-architectures to coexist and collaborate including a "reactive" layer, a "deliberative" layer and a "social" layer, along with one or more global "alarm mechanisms", a long term associative store (e.g. for answering "what if?" questions), various motive generating mechanisms, and layered perception and action mechanisms operating at different levels of abstraction. Work so far in the development of such complex notions has dealt primarily with simulated agents [36].

Being "social" implies the existence of interactive relationships. An agent capable of interactive, communicative behaviour is considered social. But, as the simple existence of two autonomous robots in the same environment forces aspects of social contact, be it direct or indirect, the necessity for a robot to have social capabilities is clear. In order for multiple robots to exhibit and maintain robust behaviours in a mutual environment necessitates a degree of social functionality.

While Dautenhahn and Christaller's [15] discussion of *body image* or *body schema* as being necessary for embodied action and cognition can be construed as an attempt to develop a more "complete" robot, it only becomes both valid and feasible within a social context.

The current state of the art in the realisation of high-level social behaviour has not extended far from a conceptual interpretation and understanding with Newell's *Social Level*, and Jennings *et al.*'s *Principle of Social Rationality*. Given the limitations of robotic hardware systems that have until recently dictated the extent to which complex control methodologies could be realised, no comprehensive research has been undertaken to date on the implications of social embodiment to the robotic domain. A wealth of anthropomorphic social analogies in the pursuit of the intelligent robot has therefore not been exploited.

References:

[1] Balch, T., Arkin, R.C. "Communication in reactive multiagent robotic systems", *Autonomous Agents*, 1, P1-25, 1994.

[2] Braaten, J. "Towards a Feminist Reassessment of Intellectual Virtue" *Hypatia*, Fall 1990

[3] Brooks, R.A., "A Robust Layered Control System for a Mobile Robot", *IEEE Jour. Rob. and Autom.*, 2(1) 1986

[4] Brothers, L. "A biological perspective on empathy". *American Journal of Psychiatry*, 146, 10-19. 1989

[5] Brooks, R.A., "Integrated Systems Based on Behaviors", *SIGART Bulletin*, Vol. 2, No. 4, August 1991, pp. 46--50.

[6] Brooks, R.A., "Intelligence Without Representation", *Artificial Intelligence Journal* (47), 1991, pp. 139—159

[7] Barnes, A., Thagard, P. "Empathy and Analogy", *Dialogue: Canadian Philosophical Review*, 1997

[8] Barnes, A., Thagard, P., "Emotional Decisions", *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Erlbaum, 426-429.

[9] Byrne, R.W. "A formal notation to aid analysis of complex behaviour: understanding the tactical deception of primates", *Behaviour* 127 (3-4) 231 – 246, 1993

[10] Clark, A. *Being There: Putting Brain, Body, and World Together Again*. MIT Press. 1997

[11] Cheney, D.L., Seyfarth, R.M., "Précis of how monkeys see the world", *Behavioural and Brain Sciences*, 15, P135-182, 1992

[12] Damasio, A. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam's Sons, 1994.

[13] Darwin, C. **"**The Expression of Emotion in Man and Animals**" (**1872) *in The Works of Charles Darwin*, ed. by Paul H. Barrett and R.B. Freeman, 29 vol. (1987-89).

[14] Dautenhahn, K. "Grounding Agent Sociality: The Social World is its Own Best Model", *From Agent Theory to Agent Implementation symposium in Proceedings of 14th European Meeting on Cybernetics and Systems Research*, 1998

[15] Dautenhahn, K., Christaller, T., "Remembering, rehearsal and empathy – Towards a social and embodied cognitive psycology for artifacts", *Proc. AISB-95 Workshop "Reaching for Mind: Foundations of Cognitive Science"*, 1995

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

[17] Descartes, R., *Discourse on Method and Meditations on First Philosophy*, Indianapolis/Cambridge Hackett Publishing 1993; 3rd edition

[18] Donald, B. R., Jennings, J. Rus, D., "Information Invariants for Distributed Manipulation" *in First Workshop on the Algorithmic Foundations of Robotics*, A. K. Peters, Boston, MA. ed. R. Wilson and J.-C.Latombe,1994.

[19] Dudek, G., Jenkins, Milios, E., M., Wilkes, D. "Experiments in Sensing and Communication for Robot Convoy Navigation" *in Proceedings IEEE International Conference on Intelligent Robots and Systems (IROS)*, 1995.

[20] H. Dreyfus. *What Computers Can't Do*. Harper, 1979.

[21] Dreyfus, H.L , *Being-In-The-World : A Commentary on Heidegger's Being and Time*, MIT Press, 1991

[22] Duffy, B.R., Rooney, C.F.B., O'Hare, G.M.P., O'Donoghue, R.P.S. "What is a Social Robot?" *10th Irish Conference on Artificial Intelligence & Cognitive Science*, 1-3 Sept., 1999 University College Cork, Ireland

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

[24] *FIPA 97 Specification: Agent Communication Language. Foundation for Intelligent Physical Agents*, 1997, http://drogo.cselt.stet.it /ufv/leonardo/fipa/spec/

[25] Gardner, H. *Frames of Mind: The theory of multiple intelligences*, New York, 1983: BasicBooks. BasicBooks Paperback, 1985.

[26] Georgeff, M.P., Ingrand, F.F., "Decision-making in an embedded reasoning system", In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 972-978, Detroit, MI., 1989

[27] Goleman, D., *Emotional Intelligence*, Bantam Books, 1997

[28] Jennings, N.R., Campos J.R., "Towards a Social Level Characterisation of Socially Responsible Agents", *IEEE Proceedings on Software Engineering*, 144 (1), 1997, 11-25.

[29] Williamsen, Kaaren "Emotions and Social Intelligence: Jane Braaten and Antonio Damasio", *Gustavus Adolphus College*, March 1995

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

[31] Kube, C.R., Zhang, H. "Collective Robotics: From Social Insects to Robots" *Adaptive Behavior*, 2(2):189-219, 1993.

[32] Lakoff, G., *Women, Fire, and Dangerous Things.* University of Chicago Press, 1987.

[33] Langton, C.G.. "Artificial Life." In C. G. Langton, (ed). *Artificial Life*, Volume VI of SFI Studies in the Sciences of Complexity, pages 1-47, Addison-Wesley, Redwood City, CA, 1989.

[34] Labrou, Y., Finin, T., "A Semantics Approach for KQML- A General Purpose Communication Language for Software Agents". *Third International Conference on Information and Knowledge Management (CIKM'94)*, November 1994

[35] Lakoff G. and Johnson. M., *Metaphors We Live By*. University of Chicago Press, 1980.

[36] Logan, B. Sloman, A., "Cognition and affect: Architectures and tools", *Proceedings of the Second International Conference on Autonomous Agents (Agents '98)*, ACM Press, 1998, pp 471—472, 1998

[37] Lucarini, G., Varoli, M., Cerutti, R., Sandini, G. "Cellular Robotics: Simulation and HW Implementation", *Proceedings of the 1993 IEEE International Conference on Robotics and Automation*, Atlanta GA, May 1993, pp III-846-852.

[38] McCulloch, W., Pitts, W. "A Logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Geophysics* 5, P115-133, 1943

[39] Mayer, J.D. & Salovey, P., "The intelligence of emotional intelligence". *Intelligence*, 17, 433-442. 1993

[40] Maturana, H.R., Varela, F.J., *"Autopoiesis and cognition – The realization of the living*", D. Reidel Publishing, Dordrecht, Holland, 1980.

[41] Neumann, J.von. *The Theory of Self-Reproducing Automata*. University of Illinois Press, Illinois, 1966. Edited and completed by A.W. Burks.

[42] Newell, A., "The Knowledge Level", *Artificial Intelligence*, 18, P87-127.

[43] Newell, A., *"Unified theories of cognition"*, Harvard University Press, 1990

[44] Olson, E.T., "The ontological basis of strong artificial life", *Artificial Life* 3, P29-39, 1997

[45] Pfeifer, Rolf, "Embodied System Life", *Proc. of the International Symposium on System Life*, Tokyo, July, 1998

[46] Pfungst, O., "Clever Hans (The Horse of Mr. von Osten): A Contribution to Experimental Animal and Human Psychology", *with an Introduction by Prof. C. Stumpf. Translated by C. L. Rahn. Reissued with an introductory essay by Robert Rosenthal.* New York: Holt, Rinehart, and Winston, 1965.

[47] Pfeifer, R., and Scheier, C. (1999). *Understanding intelligence*. Cambridge, Mass.: MIT Press. (in press)

[48] Reynolds, C. W. "Flocks, Herds, and Schools: A Distributed Behavioral Model", *Computer Graphics*, 21(4), P25-34, July 1987 (SIGGRAPH '87 Conference Proceedings) P25-34, 1987

[49] Rosenschein, J.S., Genesereth, M. R., "Deals among rational agents", *In Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* pages 91-99, August 1985.

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

[51] Rooney, C.F.B., O'Donoghue, R.P.S., Duffy, B.R., O'Hare, G.M.P.,Collier, R.W. "The Social Robot Architecture: Towards Sociality in a Real World Domain" *in Proc. Towards Intelligent Mobile Robots 99 (Bristol). Tech. Report Series*, Dept. Computer Science, Manchester University, Rep. No. UMCS-99-3-1. 1999.

[52] Ruisel, I, "Social intelligence: Conception and methodological problems". *Studia Psychologica*, 34(4-5), 281-296, 1992

[53] Sloman, A., Croucher, M., "Why robots will have emotions", *Proceedings International Joint Conference on Artificial Intelligence*, University of British Columbia, Vancouver, 1981

[54] Schulz, T.R. "From agency to intention: a rule-based, computational approach", *in Natural theories of mind: evolution, development and simulation of everyday mindreading*, A. Whiten, ed., Blackwell, Oxford,1991

[55] Searle, J., *Mind, Language and Society: Doing philosophy in the real world* Weidenfeld & Nicolson, London, 1999

[56] Sipper, M. "An Introduction To Artificial Life", *Explorations in Artificial Life (special issue of AI Expert)*, p4-8, Miller Freeman, September, 1995

[57] Sloman, A., "Towards a grammar of emotions", *New Universities Quarterly*, Vol.36, No.3, 1982.
[58] Salovey, P., Mayer, J.D. (1990). "Emotional intelligence". *Imagination, Cognition, and Personality*, 9(1990), 185-211.
[59] Schmidt, C.F., Marsella, S.C. "Planning and plan recognition from a computational point of view", *in Natural theories of mind: evolution, development and simulation of everyday mindreading*, A. Whiten, ed., Blackwell, Oxford, 1991
[60] Sterelny, K., *The Representational Theory of Mind: An Introduction*, Oxford: Blackwell, 1990
[61] Steels, L., "Building Agents with Autonomous Behaviour Systems", The `artificial life' route to `artificial intelligence'. *Building situated embodied agents.* Lawrence Erlbaum Associates, New Haven. 1994
[62] Sharkey, N., Zeimke, T., "A consideration of the biological and psychological foundations of autonomous robotics", *Connection Science*, 10, P361-391, 1998.
[63] Sharkey, N., Zeimke, T., "Life, mind and robots: The ins and outs of embodied cognition", *Symbolic and Neural Net Hybrids*, S. Wermter & R. Sun (eds), MIT Press, 2000
[64] Thagard, P., *"Mind, Introduction to Cognitive Science*", MIT Press, 1996
[65] Thorndike, E.L. "Intelligence and its uses", *Harper's Magazine*, 140, 227-235, 1920
[66] Thagard, P., Millgram, E., "Inference to the Best Plan: A Coherence Theory of Decision", *In A. Ram & D. B. Leake (Eds.), Goal-driven learning:* (pp. 439-454). Cambridge, MA: MIT Press. 1995
[67] Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem" P*roc. London Math. Soc. Ser.* 2 42, 230-265, 1937.
[68] Turing, A. M. "Computing machinery and intelligence", *Mind* Vol.59, P433-460
[69] Watson, M. & Greer, S., "Development of a questionnaire measure of emotional control", *Journal of Psychosomatic Research*, 27(4), 299-305, 1983
[70] Wispe, L. "History of the Concept of Empathy", *in N.Eisenberg & J.Strayer (eds.), Empathy and Its Development.* Cambridge: Cambridge University Press, pp. 17- 37, 1987
[71] Wispe, L., *The Psychology of Sympathy.* New York, Plenum Press (p80) 1991
[72] Worden, R.P., "Primate Social Intelligence", *Cognitive Science*, (20), P579-616, 1996
[73] Wright, I, Sloman, A., Beaudoin, L. "Towards a design-based analysis of emotional episodes", *Philosophy, Psychiatry and Psycology*, 1995.