

Robots Social Embodiment in Autonomous Mobile Robotics

Brian Duffy

University College Dublin
Department of Computer Science
brd@medialabeurope.org

Abstract: This work aims at demonstrating the inherent advantages of embracing a strong notion of social embodiment in designing a real-world robot control architecture with explicit “intelligent” social behaviour between a collective of robots. It develops the current thinking on embodiment beyond the physical by demonstrating the importance of social embodiment. A social framework develops the fundamental social attributes found when more than one robot co-inhabit a physical space. The social metaphors of identity, character, stereotypes and roles are presented and implemented within a real-world social robot paradigm in order to facilitate the realisation of explicit social goals.

Keywords: mobile robotics, artificial intelligence, agent architectures, intelligent systems design.

1. Introduction

To date, a fundamental facet of embodiment has, on the whole, 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 its 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 robot’s motions, whether abstract notions of a moving obstacle or its clear distinction of another individual robot, influences the observing robot’s behaviour. The social implications of two robots coexisting in an environment add another dimension to the complexity of each robot’s perception of that 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 physical embodiment and the “social” connotations of one robot’s influence over another becomes apparent. While this abstract notion of “communication” is not necessarily explicit in either its intention or application, it does not constitute a “degree

of sociality” in the sense of a structured interaction. Work on emergent robotic systems have embraced such inter-robot dynamics without explicit social representations or communication (Mitsumoto et al, 1995; Fukuda et al, 1989; Cai et al, 1995; Holland & Deneuborg, 2000; see Cao et al, 1997 for an overview). These systems have been characterised as exhibiting “emergent intelligence” and constitute many simple, independent, individual robot agents acting according to a purely local perception of their environment and following a set of simple rules to actuate local environmental changes.

When looking at explicit social interaction between two or more robots capable of engaging in some degree of communication, a whole new set of problems arise. Worden (Worden, 1993) discusses the structure of such social domains 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.

While some points are open to discussion, an idea of the issues to be addressed emerges. The following sections start by briefly reviewing work to date on the development of strong social functionality in multi-agent systems research. The required social functionality for a team of autonomous mobile robots required to undertake explicit social tasks in real-world problem domains is then presented. Section 6 then demonstrates how such social features are realised on a team of socially capable robots..

2. Sociality in Multi-Agent Systems Information

Multi-Agent Systems research has looked to address prevalent issues in the social interaction of a team of agents and as such provides us with a rich source of ideas. The inherent notion of multiple agents required to interact in a coherent reproducible manner is a core issue. Newell's unified theory of cognition (Newell, 1990) 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 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 and Campos (1997) et al. introduce the Social Level Hypothesis 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) (Newell, 1982). 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.

Jennings and Campos discuss the Social Level from the context of social responsibilities and leads to the formulation of the Principle of Social Rationality.

2.1. Principle of Rationality and of 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 (1982) 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. (Jennings & Campos, 1997) 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"

Justifications for the extension of Newell's original proposal to the Principle of Social Rationality are based on the balance between the individual benefit between member's 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, an explicitly social community of agents can therefore exist. In order to achieve these attributes, the "tools" facilitating social interaction such as communication are required.

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, limited comprehensive research has been undertaken to date on the implications of social embodiment to the robotic domain (Duffy, 2000). A wealth of anthropomorphic social analogies in the pursuit of the intelligent robot has therefore had limited exploitation.

3. Social Power

In a paper entitled "Social Power: A point missed in Multi-Agent, DAI and HCI" (DAI – Distributed Artificial Intelligence, HCI – Human Computer Interface), Castelfranchi (1990) suggests that there has been a "serious lack of realism in Multi-Agent and interaction studies" where "sociality or the agents is merely postulated rather than explained in terms of their dependence".

In addressing this issue, Castelfranchi proposes a distinction between Distributed Artificial Intelligence and what he terms the "Social Simulation Approach", i.e. the difference between the research concerned with intelligence, problem solving and system architecture where society is "used as a metaphor", and, on the other hand, work that deals specifically with social interaction. This corresponds to the distinction between the two fundamental issues present in social scenarios:

The social empowerment of a robot: The attributes required by an agent in order to consider it socially capable.

Task decomposition problem: The issue of how a global task can be decomposed and completed by a set of socially capable agents.

Castelfranchi uses alternative terminology and refers to the social empowerment of a robot as the Sociality Problem, and the task decomposition issue as the Adoption Problem and deals primarily with the second from the perspective of “social power”, i.e. the “power” one agent has over another in a social environment (Castelfranchi, 1990). Conte et al. within the Social Behaviour Simulation Project also maintains a distinction between goal adoption and cooperation (Conte, Miceli & Castelfranchi, 1991).

This paper proposes that the two cannot be isolated and postulates that in order to develop an artificially intelligent physically situated robotic entity, embracing a strong notion of social embodiment is a necessary criterion. A robot must have both the capabilities to be social in conjunction with its abilities to solve social problems. While this is not a new notion, it has not been developed in the context of robotics and artificial intelligence as an all-encompassing coherent approach.

As highlighted by Castelfranchi, the fundamental issue appears to be the void between the social goal and the social robot, i.e. breaking down the global task into some set of subtasks to be completed and the social empowerment of a robot. Multi-Agent Systems have proposed numerous strategies and models for the task decomposition problem (Sekiyama & Fukuda, 1999; Durfee & Lesser, 1987; Lesser et al, 1998; Cohen & Levesque, 1997; Grosz, 1996; Stone & Veloso, 1999) but little has been done in developing true agent sociality in multi-agent systems. When the issue is the social interaction of a number of autonomous robot entities, serious considerations arise, i.e. resource bounding, incorrect perceptions and numerous other attributes inherent with real world applications (see Duffy, 2000 for a discussion). The next sections therefore aim at taking intentional multi-robot systems a stage further by developing the “social robot”.

4. Towards Sociality in Autonomous Mobile Robotics

Panzarasa et al. (1999) present a conceptual model for representing the inherently social implications of multi-agent systems based on an agent’s individual mental states. They propose the use of social mental shaping via roles and social relationships and gives examples of some of the ways in which social relationships can drive an agent’s behaviour by influencing its mental state:

- Authority: hierarchical social status.
- Helping disposition: i.e. altruism.
- Trust: based on the confidence an agent has for another.
- Persuasion: i.e. through a process of argumentation.

In order to achieve such relatively high-level social functionality, a set of “tools” is required. While Panzarasa et al. propose that roles and social relationships “may complement and augment bare individual mental attitudes”, this paper proposes that the concept of identity is necessary in conjunction with character, stereotypes and roles in order to achieve social relationships between robots.

Kinny et al. (Kinny et al, 1996) propose the notion of the “internal” and “external” aspects of an agent where the internal features of the agent comprise its beliefs, desires, and intentions, while the external features relate to features of the social group, i.e. the roles and relationships within the system. An important distinction arises when considering embodied systems over software based virtual environments. In this work, the notion of internal and external relates to the attributes of a single embodied agent or robot analogous to Shoham’s notion of capabilities in multi-agent systems (Shoham, 1993). The internal attributes of the robot are analogous to Kinny et al.’s internal features. While Kinny et al.’s external aspect relates to the social interaction, i.e. the services an agent provides, its interactions, and the syntax and semantics for the communication between agents, this notion is here developed further with the addition of stereotypes, roles and characters as the social features.

There is an important difference between virtual or software-based agents and embedded systems. In real-world robotics there is an extra dimension of complexity of physical manifestation. In order to encompass the added complexity of dealing with embodied agents (both physically and socially), the original internal and external features of an agent proposed by Kinny et al. are insufficient, as they do not address the added complexity of the physical environment. A discussion of strong physical and social embodiment can be found in (Duffy, 2000; Duffy & Joue, 2000). It follows that in order to address the issue of embodiment, there are two distinct robot attributes that are local and particular to each robot within the social system:

- Internal Attributes: Beliefs, desires, intentions (based on Rao & Georgeff, 1989), i.e. name, character, mental capabilities, the robot’s knowledge of self, experiences, a priori and learned knowledge, processing capabilities (i.e. algorithms, DSP)
- External Attributes: the physical presence of the agent in an environment; its actuator and preceptor capabilities (i.e. a robot equipped with extra sensory equipment compared to another), the physical features of the robot, i.e. physical dimensions.

And one global system attribute which subsumes the social functionality of the collective of robots:

- Social Attributes: Identity, character, stereotype, roles.

The Social Attributes are more abstract social features that exist to facilitate the interaction between robots. While some pertain to the robot itself, they nevertheless constitute attributes existing in the social system that are necessary for the social functionality of the system.

These attributes are developed in greater detail in the following sections.

5. Social Functionality

Complexity issues arise when the complete problem encompasses multiple agents, with different social tasks, and are embodied in a complex real world scenario. This necessitates the development of suitable formalisms to facilitate the resolution of allocating social tasks without becoming overwhelmed by the added complexity of the robot's environment, both physical and social. The formalisms are therefore developed from each robot's perspective, not from a complete systems perspective. It is the fundamental notion of embodiment that necessitates this approach over that proposed by Kinny et al (1996).

5.1. Identity

When social interaction exists, each element of the social group must be able to be differentiated from others. The robots require a sense of themselves as distinct and autonomous individuals obliged to interact with others in a social environment, i.e. they require an *identity*. Identity refers to the property appropriate or peculiar to an individual, that quality which distinguishes one person or thing from another. Identity and the "self" have emerged as a central focus of theory and research in many domains of social and behavioural science (Baumeister, 1989). *"The large volume of empirical research on the self has convinced most social and behavioural scientists that the self is real, and that no science of the human experience is complete without accounting for it"* (Hoyle, 1989).

Suppose robot R_i is a part of a social group S . A social group must involve more than one robot. That is,

$$\exists R_i, R_j \in S \mid R_i \neq R_j \wedge S \subseteq R$$

where

R is the set of all possible robots

Given that this research deals primarily with four Nomad Scout II robots named Aengus, Bodan, Bui and Caer, this can be written more specifically as:

$$\{\text{Aengus, Bodan, Bui, Caer}\} \in S$$

Identity and embodiment are inherently linked. The manifestation of mental capabilities and knowledge in a physical body is the ascription of a concrete singular identity to a "brain". It is the synthesis between brain and body with the social implications of co-existing in an environment where identity is the foundation stone for social interaction. The role of embodiment should not be trivialised (Sharkey & Ziemke, 1998; Clark, 1997; Duffy et al, 2000). The ascription of the physical world to primarily software based agent research existing in virtual environments can be perceived as an added degree of complexity. On the contrary, the physical world acts as the base on which social concepts are grounded. It constrains the highly complex notions of multiple identities, and accountability.

In a social environment, knowing the identity of those with whom one communicates is essential for

understanding and evaluating an interaction. Contrary to the disembodied world of virtual multi-agent communities where identity can be ambiguous (Donald, 1994), real world applications of robotic agents have identities inherently based on their physical existence. Many of the basic cues regarding personality and social role that we are accustomed to, are prevalent in the physical world, for example, statically being the body's physical construction, and dynamically being motion behaviours.

As mentioned above, the physical body provides a compelling and convenient basis for the definition of identity. Though the "brain" may be complex and mutable over time and circumstance, the body provides the stabilising anchor. There is therefore a one to one mapping between a robot and an identity. While alternate social interaction spaces, i.e. virtual communities (Donald, 1994), may facilitate multiple personas, real world social spaces are constrained by the physical embodiment of the "brain" in a single robot platform. Identity cues are primarily based on the physical attributes of a robot, i.e. its actuator and perceptor capabilities. A complete concept of *identity* therefore constitutes the set of internal and external attributes associated with any given robot based on introspection of its physical and "mental" states and capabilities. An identity can develop to a certain extent with its acquisition of knowledge and experiences over time. It is an individual attribute of each robot and constitutes the robot's true perspective of self. Identity is fundamentally based on *what* the robot is and as such comprises its *internal and external attributes*.

Smithers (1995) discusses *identity* from the perspective of an observer where an entity (robot) is "*required to have the means to introduce additional dynamical properties in the interactions with its environment*". An example is given where "*if the environment forces inducing motion change or go to zero, the behaviour of the entity-environment system may change, and change in such a way the original entity can no longer be picked out from the general goings on in its environment: its identity may be lost*". While this perspective is fundamentally based on classical approaches to robotics where the controller is simply placed in its environment, the social implications of multiple robots force a true notion of identity. This true notion will therefore differentiate a particular robot from its environment and is primarily independent of any particular behaviour that it may or may not exhibit.

Following from this, a definition of *robot identity* can be derived:

Robot identity:

differentiates a system from any other and preserves its distinction from its environment regardless if it has any observable influence on, or actuation with, its environment

This approach subsumes the mere fact that the entity (robot) physically exists in the environment and has therefore, by its simple physical presence, changed its environment, which induces a form of physical identity.

Each robot possessing a particular identity, and therefore developing and exhibiting that particular robot's capabilities in a social environment (and promoting it to specialise in the attributes relevant to its identity), presents a very interesting problem. It is proposed that each autonomous robot has an individual *identity*, which will facilitate the development of the "social" aspects of multi-agent interaction between such autonomous agents. (Note on Social Identity: Members of social categories and groups may have a sense of "social identity" relative to a global task that dictates the formation of the group. This is not discussed here).

The identity of a robot is made up of its internal and external attributes. Examples of *internal-attributes* include: language, sensory processing capabilities (i.e. vision, filtering, or smoothing algorithms), social knowledge, and physical environmental knowledge. Examples of *external-attributes* include: camera, sonar, wheels and motors, bumper, i.e. the physical features of the robot.

The *identity* of R_a or $I(R_a)$ is therefore founded on the union of both its particular *external attributes* EA (i.e. sensors, actuators, dimensions) and *internal attributes* IA (i.e. knowledge of its social and physical environment, processing capabilities, plan library):

$$I(R_a) = EA(R_a) \cup IA(R_a),$$

where

$$\begin{aligned} \exists ea_i, ea_j \in EA(R_a) \mid ea_i \neq ea_j \text{ and} \\ \exists ia_i, ia_j \in IA(R_a) \mid ia_i \neq ia_j \end{aligned}$$

Thus

$$\begin{aligned} EA &= \{ea_1, ea_2, ea_3, \dots, ea_n\} \text{ and} \\ IA &= \{ia_1, ia_2, ia_3, \dots, ia_m\} \end{aligned}$$

which gives:

$$I(R_a) = \{ea_1, \dots, ea_n, ia_1, \dots, ia_m\}$$

Shoham (1993) assumes that capabilities are fixed. Intuitively, the use of internal and external attributes dictate that some are static and others are dynamic.

Here it is understood that the external features of the robot (i.e. sensors, actuators) are static, while the internal features are indicative of the knowledge, experience, and processing capabilities (i.e. smoothing or filtering algorithms) that a robot has either obtained over time or

has been initially provided with, and are therefore dynamic.

A robot's knowledge of its attributes therefore allows a sufficient degree of introspection to facilitate and maintain the development of social relationships between groups of robot entities. When a robot is "aware" of itself, it can explicitly communicate knowledge of self to others in its social environment.

As the *identity* of each robot is governed by a combination of its internal and external attributes, any changes in its internal attributes will therefore have an effect on its identity. The *identity* of the robot can therefore develop over time.

5.1. Character

An important criterion for the development of a social environment is the attribution of mental states to the internal representation one robot has for another. Robot R_1 must develop its own beliefs about the internal state of robot R_2 in order to develop a higher-level social interaction with the robot. This leads to the development of complex dynamic social interactions in a real world environment and a corresponding added complexity to the system.

The use of the terms "identity" and "character" in mobile robotics, or more generally embedded systems, is so far limited to non-existent (see Duffy, 2000 for a review). This work highlights that the distinction between a robot's identity and its perceived character should not be confused. An illustration is where two physically similar people (i.e. identical twins) have two different identities. While having two similar bodies, each can display distinctly different behaviours, as can manifest in different trains of thought, views, beliefs, ideals and temperament, experiences and memories. This highlights how someone's identity is influenced not only by its appearance but also by its behaviour. Likewise, a robot's character depends on observation. The identity of robot R_1 that robot R_2 perceives may be different from that perceived by robot R_3 .

In this work, *character* is the *perceived* identity of one robot by another. Character is based on initial stereotype bootstrapping (see Section 5.3), communicated knowledge, and/or experience in working with that robot over time. It corresponds to another robot's "interpretation" of one robot's identity based on the knowledge it has about that robot.

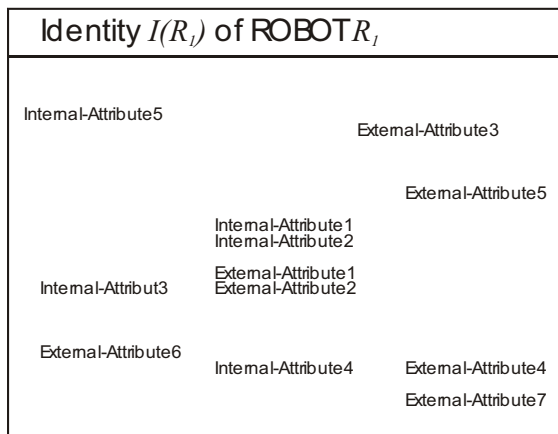


Fig. 1: Identity is the set of a robots internal and external attributes

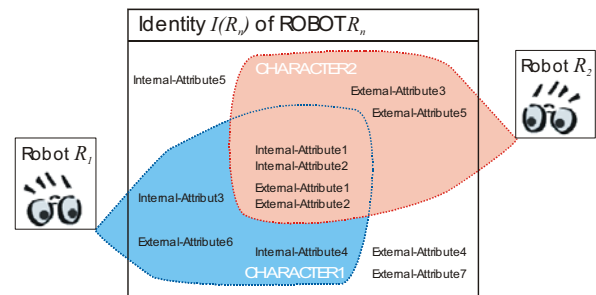


Fig. 2: The perceived identity (or *character*) of robot R_n by two other robots R_i and R_j

Character: The combination of perceived features or qualities that distinguishes one entity from another in that entity's social envelope.

Character deals with the fundamental attributes an agent or robot is *perceived* by others in its social environment to have, its capabilities albeit physical or mental. *Character* represents a subset of the list of internal and external attributes of each robot. The character of robot R_2 has of the robot R_1 at any time t may or may not be the same as the character that robot R_3 has of the robot R_1 (as shown in fig. 2).

Each robot builds a list of representations or *characters* for each other robot that exists in its social environment (fig. 3). The character of robot R_j as perceived by robot R_i at time t_n is:

$$C(R_j, R_i t_n)$$

Ideally there is a one-to-one mapping between a perceived character and that particular robot's identity, where each robot would know all other robots "completely", i.e. it would have complete knowledge of all other robots in its social environment. This would in fact rarely happen and would also be unnecessarily complex. Character is a subset of the total set of internal and external attributes that comprise identity:

$$C(R_j, R_i t_n) \subseteq I(R_i)$$

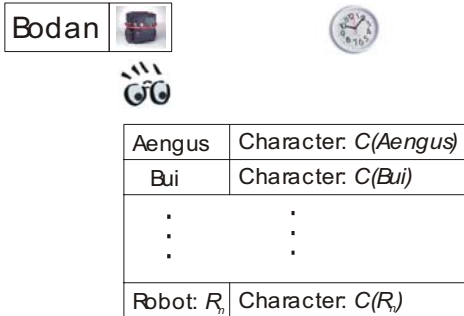


Fig. 3: Robot Bodan's "view" of the characters of robots Aengus, Bodan & etc at time t

The use of heterogeneous robots also facilitates unique perceived characters in social environments. A robot equipped with a vision system may become the "eyes" of the group where the physical construction of the robot plays an important part in defining its function in the social group. The "role" a robot plays is discussed in greater detail in Section 5.4.1.

The use of *characters* therefore facilitates the building of internal socially-grounded representations for all other robots in a particular robot's social environment, a fundamental component of high-level social interaction.

5.2. Stereotypical Representations

While a robot's *character* is transient over time, this *perceived* identity of another should be strongly founded on some fundamental set of internal and external attributes that describe that robot. This ensures that the representations of another robot's identity remain valid

for different robots. This is achieved by the use of stereotypical representations, or *stereotypes*. A fixed subset of internal and external attributes comprises the *stereotype* with which each robot is associated.

While character representations are initially independent of a global task or set of subtasks to be undertaken, the notion of stereotypical representations is proposed in order to bootstrap the initial stages of social interactions and reduce the complexity of maintaining such internal representations. The perceived character of a robot is fundamentally based on a set of stereotypical representations available to each robot and developed through communication, collaboration and experience. At time t_0 when robot R_1 meets robot R_2 , such stereotypical representations are used to facilitate the "introduction" procedures between the two robots, i.e. the initial communication to ascertain who and what the other robot is. Each robot is equipped with a finite set of stereotypical representations. These stereotypes "moderate" the character representations that one robot builds of another.

Each robot knows its name, its associated stereotype and the attributed internal and external attributes associated with this stereotype. This allows for the introduction of a robot with a new stereotype to the group. As it knows all details pertinent to itself, it can communicate this to others in its social environment who will then learn the new stereotype and corresponding internal and external attributes.

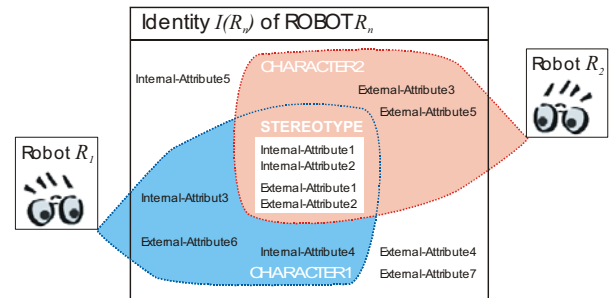


Fig. 4: A defined subset of internal and external attributes constitutes the stereotype that robot R_n is associated with.

Each robot has one stereotype associated with it. The stereotype of the robot R_a is given by:

$$St(R_a)$$

There are a fixed number of stereotypes in the social environment:

$$Stereotype List = \{St_1, St_2, \dots, St_m\}$$

A stereotype comprises a defined finite list of *internal attributes* and *external attributes*:

$$St_x = \{ea_1, \dots, ea_n, ia_1, \dots, ia_m\}$$

The stereotype of robot R_a is a subset of its *character* as perceived by robot R_β at time t , which in turn is a subset of its identity:

$$St(R_a) \subset C(R_a, R_\beta t) \subset I(R_a)$$

Each robot in the social group has knowledge of the possible stereotypes that may exist in its social environment and all details pertaining to each stereotype in that list. A robot therefore, knowing the stereotype

associated to a particular robot it encounters, will also know all the internal and external attributes associated with the stereotype of that robot. Each robot in the social group may see differing characters of a particular robot, but all are fundamentally based on that robot's *stereotype* where the stereotype is the foundation stone for all other robot's perception of its identity (see fig. 4). An example is where robot R_1 is equipped with a special pallet-carrying device and its corresponding stereotype association is *Pallet_Carrier*. When this robot becomes a member of a social group, all other robots build an internal representation of this robot. They are aware of the basic internal and external attributes associated with a *Pallet_Carrier* (e.g. arms, sturdy load carrying chassis, lift control algorithms, etc). While the identity of the pallet-carrying robot R_1 comprises many (including some dynamic) attributes, it is impossible for all other robots in the social group to know all these attributes. They build representations of robot R_1 based on their perceived identity of robot R_1 , i.e. character, with the initial knowledge that it is a *Pallet_Carrier* (its associated stereotype). Their character representations are therefore bootstrapped with the use of a stereotype. While a number of robots may have the same stereotype, i.e. "looker" for those equipped with vision systems, their character models of each robot will develop over time and can develop into quite different *characters* (whilst always based around their stereotype associations as shown in fig. 4).

The introductory contact between two robots is therefore initially bootstrapped by the stereotype that each robot is associated with, which facilitates the development of the internal representation of other robots in the social environment and their corresponding functionality regarding tasks to be performed by the group. A degree of recognition is preferable to self-identification as the conversation can use simple "words", then jump to a more complex level. If there is no recognition (or categorisation based on stereotype), then robot R_1 must find out who/what robot R_2 is, and must specifically request such information.

The personal, physical and social knowledge presented in the previous sections is stored in the form of beliefs with each robot's belief set, a feature of BDI-based architectures (Rao & Georgeff, 1989; Jennings, 1993; O'Hare & Jennings, 1996; Duffy *et al*, 1999). Examples of this are shown in Section 7.

The use of stereotypes is indicative of exactly *how* another robot can help in completing a required global social task. This facilitates the breakdown of complex collaborative tasks into individual robot allocated subtasks by reinforcing the communication of elements of one robot's *identity* to another. The stereotype list of robots in the collective therefore provides a basis for subtask allocation and is discussed in the following section.

5.4 The Global Task

The social requirements of a social task dictate that the task must be allocated to a collective of robots. In order

to have a social group of collaborating robots perform such a global social task, the subdivision of a social task into suitable subtasks that can be performed by the co-operating collective of robots is required. These subtasks must be allocated relative to each robot's abilities to perform each subtask thereby dictating that all robots should be altruistic in nature. Commitments are then required by all robots to collectively work towards the global objective or goal.

When the global task has been decomposed into a set of subtasks, the issue is how to allocate these subtasks to appropriate robots within the social group. The notion of "role" is introduced to facilitate this subtask allocation.

The task decomposition issue as a topic of research is beyond the scope of the work presented and is therefore not discussed in great detail (see Haddadi, 1995 for a formal description). This complex problem constitutes a major field of research in both multi-agent systems research and artificial intelligence in general. As the allocation of a social task to a collective of socially competent robots necessitates both task decomposition and the development of existing technologies to encompass the social implications of such a problem, the notion of "role" is proposed and discussed in suitable detail.

5.4.1 Role

A robot must undertake a *role* in order to fulfil its required work within a social group so that the group can achieve a specific social task. Kinny *et al.* (1996) similarly discuss *role* in the context of an agent's participation in the social group where the agents are broken up into groups based upon their particular roles and relationships within the system. The *role* a robot must undertake can be defined as:

Role: The characteristic and expected behaviour of an individual with regard to a particular social goal or task within a social collective of individuals.

Here the term "expected" is used in the context of the required social behaviour the robot has to undertake or perform.

A role is primarily task driven, i.e. the role the robot must undertake to complete a task is analogous to the role an actor must undertake in a performance so that the play will be performed.

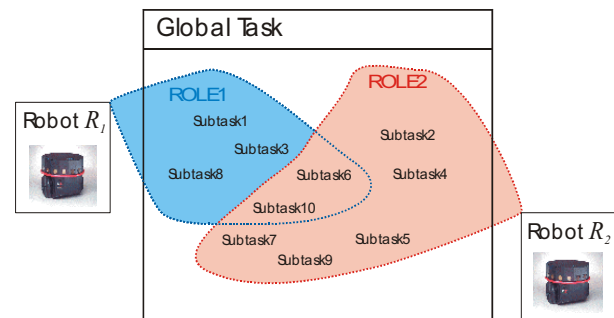


Fig. 5: The roles that robot R_1 and robot R_2 are required to undertake in the realisation of the global task.

While the mapping of a role to a robot is dependent on the robot's capabilities, it may only require a subset of its capabilities to complete. A role is task orientated and will therefore change with different tasks, different times, and different social colleagues. Each robot will therefore have a *role* to fulfil within a social community of robots, i.e. it must complete its assigned subset of subtasks as shown in fig. 5.

The global task is decomposed into subtasks, which are grouped together as *roles* in order to be allocated to the appropriate robots in the social group based on what each robot can do. This role allocation is based on the stereotypes existing in the group. From the previous example, the `Pallet_Carrier` robot (`Pallet_Carrier` being its stereotype association) can undertake those roles pertinent to such capabilities. For example, it can carry pallets of raw materials from one robot to another.

A *role* therefore corresponds to a subset of the social plan of tasks to be performed by an individual robot, and the corresponding behaviours to achieve these tasks. A task constitutes undertaking to perform a plan of behaviours. Formally, a plan p denotes a plan expression in U_p , and a plan expression consists of actions and the "sub-plan" expressions (following from the formalisms presented in Haddadi, 1995). Developing from Haddadi's work on actions and plans, the use of *roles* corresponds to sub-plans with the global task being the main plan. The semantics for the action formula is denoted by: $(\text{Has_Plan } R \ p \ \psi)$ where robot R has a plan p in its plan library to achieve the state of affairs ψ .

$$\langle M, V, w_{t_0} \rangle \models (\text{Has_Plan } R \ p \ \psi)$$

$$\text{iff } [[P]] \in \Pi([R])$$

and

$$\langle M, V, w_{t_0} \rangle \models A \ \square \ ((\text{Succeeded } R \ p) \Rightarrow \psi)$$

where

R - a robot

M - a model

V - function mapping model to world (variable assignment)

W - a world at time t_0

\square - always

Π - a plan library

Ψ - a plan

In recapping, robot R will carry out some plan to achieve ψ :

$$(\text{Achieves } R \ p \ \psi) \stackrel{\text{def}}{=} (\text{Has_Plan } R \ p \ \psi) \wedge (\text{Does } R \ p)$$

Note that this does not guarantee that after p is executed, ψ will be successfully achieved. Successfully achievement is defined as:

$$(\text{Achieved } R \ p \ \psi) \stackrel{\text{def}}{=} (\text{Has_Plan } R \ p \ \psi) \wedge (\text{Succeeded } R \ p)$$

By extending Haddadi's use of the term "Contributes" to denote how a plan p contributes to achieving the state of affairs ϕ by first achieving the state of affairs ψ , where the state of affairs ϕ represents the social plan to be undertaken by the social group:

$$(\text{Contributes } p \ \phi \ \psi) \stackrel{\text{def}}{=} A \ \square$$

$$((\text{Succeeded } p) \Rightarrow \phi) \wedge E(\psi \ \mathcal{B} \ \phi)$$

where

E - optionally

\mathcal{B} - a belief

This constitutes the *role* that a robot plays in the social group. It is the contribution a robot makes to the completion of the social task. The social task therefore results in a social plan of specific robot plans that are to be allocated to members of the social group, where *role* allocation is based on each member's stereotypes (i.e. what they are capable of achieving).

In order to delegate the tasks between robots, the following definitions are required:

- Robot R_i will engage in cooperation with robot R_j with respect to achieving a goal state ϕ , such that R_j achieves at least ψ , if:
 - R_i has ϕ as a goal
 - R_i is not willing to achieve ϕ individually
 - R_i believes that one way of achieving ϕ is first to achieve ψ , and has ψ as a subgoal
 - R_i has as a goal that R_j achieves at least ψ
 - R_i does not believe that R_j would disagree with these terms

Haddadi refers to this delegation as the `potential_for_commitment` or PfC and continues in discussing pre-commitment and commitment with respect to R_j achieving ϕ and expresses it formally as:

$$(\text{PfC}_d \ R_i \ R_j \ \phi \ \psi) \stackrel{\text{def}}{=} (\text{GOAL } R_i \ \diamond \phi) \wedge \neg(\text{Willing } R_j \ \phi) \wedge (\text{GOAL } R_i \ E(\psi \ \mathcal{B} \ \phi)) \wedge$$

$$(\text{GOAL } R_i \ E(\text{Achieves } R_j \ \psi)) \wedge \neg(\text{BEL } R_i \ \neg(\text{Pre_commit}_a \ R_j \ R_j \ \psi))$$

- Robot R_i is `pre_committed` to robot R_j if R_i sees a potential for cooperation with R_j such that R_j achieves ϕ , and R_i wants that R_j achieves ϕ :

$$(\text{Pre_commit}_d \ R_i \ R_j \ \phi) \stackrel{\text{def}}{=} (\text{PfC}_d \ R_i \ R_j \ \phi) \wedge (\text{Want } R_i \ R_j \ \phi)$$

- When robot R_i is `committed` to robot R_j achieving ϕ then R_i has committed to the choice of R_j achieving ϕ , and R_i believes that R_j will eventually commit to achieving ϕ :

$$(\text{Commit}_d \ R_i \ R_j \ \phi) \stackrel{\text{def}}{=} (\text{INTEND } R_i \ E(\text{Achieves } R_j \ \phi)) \wedge (\text{BEL } R_i \ \diamond E(\text{Achieves } R_j \ \phi))$$

Haddadi's `potential_for_commitment` is practically realised in this work by the concept of robot *stereotypes* discussed in Section 5.3.

A collective of socially capable robots adopting a social task is strongly founded on their capacity to communicate and a language of communication that adheres to the following formal structure (see Rooney, 2000 and Duffy *et al*, 1999) for a description of the communication language Teanga which has the required expressive power for such BDI-based approaches):

- Robot R_i sees a `potential_for_commitment` with robot R_j with respect to its goal state ϕ through its

knowledge of R_j 's stereotype, such that R_j achieves at least ψ if

1. R_j believes that R_i has the goal that R_j achieves φ
2. R_j believes that one of the ways of achieving φ would be to first achieve ψ and has ψ as a subgoal
3. R_j has a goal to achieve ψ by carrying out some plan of actions
4. R_j does not believe that R_i would disagree with those terms

$$(\text{PFC}_d R_j R_i \varphi \psi) \stackrel{\text{def}}{=} (\text{BEL } R_j (\text{GOAL } R_i \diamond \varphi) \wedge (\text{GOAL } R_j \text{E}(\psi \mathcal{S} \varphi)) \wedge (\text{GOAL } R_j \text{E}(\text{Achieves } R_j \psi)) \wedge \neg(\text{BEL } R_j \neg(\text{Pre_commit}_a R_j R_j \psi))$$

- Robot R_j is **pre_committed** to robot R_i in achieving φ . That is, R_j sees a potential for cooperation with R_i through its knowledge of R_j 's associated stereotype such that R_j achieves φ and R_j is willing to achieve φ :

$$(\text{Pre_commit}_d R_j R_i \varphi) \stackrel{\text{def}}{=} (\text{PFC}_d R_j R_i \varphi) \wedge (\text{Willing } R_j \varphi)$$

- Robot R_j is **committed** to robot R_i in achieving φ thus R_j intends to carry out a plan that will achieve φ and R_j believes that R_i will eventually commit to the choice of R_j achieving φ :

$$(\text{Commit}_d R_j R_i \varphi) \stackrel{\text{def}}{=} (\text{INTEND } R_j \text{E}(\text{Achieves } R_j \varphi) \wedge (\text{BEL } R_i \diamond (\text{INTEND } R_i \text{E}(\text{Achieves } R_j \varphi)))$$

For further information see Haddadi (1995), and in particular for detailed formalisms on the axioms of Pre-commitments and Commitments.

The pre-condition of mutual altruism subsumes the issue of a robot being committed to another robot in undertaking a subtask and is thus a necessary initial feature of the social group.

A robot may be assigned any number of *roles*. In such cases, there is temporal ordering of such *roles*. A *role* may have any number of subtasks, in any order, with possible repetitions of subtasks. If one attempts to assign a global task to a collective of robots that do not have the capabilities to perform what is required of them, this will become apparent via the stereotypes within the collective. The stereotypes will therefore constrain the incorrect allocation and subsequent robot functionality-based failure of the task based on their capabilities.

When the agents are obliged to play different *roles* in the group, the system of multi-agents are constrained in their mutual interaction. These roles therefore impose requirements on how they are to behave and to interact with others and result in a strong co-ordination structure. This abstract notion of *roles* between the global task level and the set of elemental behaviours facilitates the global task decomposition and allocation process and provides a degree of abstraction for communication between collaborating robots.

As physical attributes (external) of a robot are assumed static and absolute (i.e. the hardware configuration) and the internal attributes are more dynamic in nature (i.e. environmental knowledge), the *robot* \Leftrightarrow *role* allocation is primarily based on a robot's physical attributes and its

associated internal attributes (i.e. a camera and vision algorithms) with the dynamic internal aspects acting as more influential when indecision exists as to whether robot R_1 or robot R_2 can be assigned a role r_j .

In a group of homogeneous robots where their physical construction is similar and as such negate the influence any external attributes have on the role allocation process, the internal attributes influences the allocation of roles. The evolution of a robot's internal states over time based on experience, learning and knowledge accumulation also can increase the importance of the *internal attributes* when task allocation is being undertaken (i.e. a robot that has learned the environmental map for the workspace concerned). The mapping function between the robot and the role it is required to perform should be able to deal with such temporal issues.

It is important to note that the decomposition of the global task into subtasks is dependent on the stereotype list, not on each robot's individual plan library. The global social task is not decomposed dependent on conditions unique to any one robot but rather the conditions prevalent to the social group as a whole.

5.4.2 The Subtasks

A subtask constitutes a plan of elemental behaviours that can be executed by the appropriate robot. Each subtask may have any number of behaviours (i.e. `follow_wall`, `avoid_obstacle`, `take_snapshot`), in any order, with possible repetitions of behaviours. These behaviours are dynamic in nature and may be over-ridden by reflex behaviours in emergency situations. Duffy (2000) develops this feature in more detail by presenting an architecture with sufficient functionality to support the concepts proposed here.

Each subtask t_i is comprised of a subset of all possible behaviours B :

$$t_i \subseteq B = \{b_1, b_2, \dots, b_n\}$$

Any given subtask may contain any behaviour in any order with possible repetitions of behaviours.

The behaviours are temporally ordered, ensuring that some behaviours are only initiated when appropriate others have been completed.

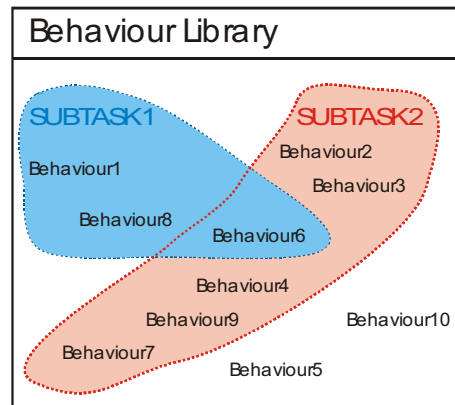


Fig. 6: Each subtask comprises a subset of behaviours that are stored in the behaviour library.

A behaviour constitutes a mapping between a sensory stimulus and actuator functionality, but may on occasion be simply either sensory information or actuator motion. All tasks are undertaken via a behavioural hierarchy (see Duffy, 2000).

5.5 The Whole Picture

The concepts of identity, character, stereotypes, and roles proposed in this work have been developed as a complete and integrated framework to facilitate the development of a social community of robots with suitable functionality to complete required social tasks.

In returning to the minimum set of concepts necessary for social behaviour and a society as defined by the Principle of Social Rationality (Jennings & Campos, 1997), the following objectives have been undertaken:

- Defining an independent primitive for each agent, i.e. its identity.
- Defining an independent primitive for one agent's model of another agent, i.e. its character.
- Description of task decomposition notions such as roles analogous to (Panzarasa Normal & Jennings, 1999; Kinny et al, 1996).
- Assessing what "capabilities" exist in the co-operative, i.e. a global list of the attributes available to achieve the global task, i.e. stereotype listing.
- A mapping between the robots and the tasks to be undertaken (i.e. which robot undertakes which role based on its stereotype association).

The system objectives address the notions of acquaintance, influence, and rights and duties in a social collective of embodied robotic entities. The first two are new approaches to multi-robot control methodologies. This work proposes that the definition and development of these primitives facilitates the development of a complex social structure between a collective of socially capable robotic entities functioning in real-world environments. This not only addresses the issue of developing from the conceptual notion of the Principle of Social Rationality, but also seeks to apply this to the real world domain.

5.5.1. The Conditions

The following set of conditions and pre-requisites are imposed on the system in order for a collective of autonomous robots to complete a global task:

- The robots are altruistic in nature: (i.e. it is assumed that all robots in the social group will undertake to perform whatever tasks are required of it so that the group will achieve an explicit social goal required of the group)
- Good faith: It is assumed that agents commit only to what they believe they are capable of. This also includes that once a robot commits to something that it will undertake to achieve it.
- The robot knows its own attributes and correspondingly its identity, i.e. knowledge of its external (physical) and internal state.

- Handshaking is initially performed by each robot with any other robots in a group and by any robot entering an existing group to all other robots in the group. It is a requirement of a new robot to "introduce" itself and its identity to all others.
- The global task must be divided into realisable subtasks and allocated to appropriate robots, equipped both internally and externally, to achieve the task.
- Each robot should have a degree of autonomy in performing its allocated task.
- Communication between robots should be possible in order to update each robot's knowledge about other robots' contribution to the global task.
- Each robot should communicate its completion of its subtasks.
- An evaluation of the result is required to ascertain that the task (both at a global and sub level) has been completed correctly.

Following from Shoham's informal guidelines (Shoham, 1993) on the change and persistence of mental attitudes over time, it is also assumed that beliefs must be adopted or learned and persist by default. That is, once a belief is adopted, this will remain until either the belief is specifically let go or a contradictory belief is adopted.

While some of the above appear trivial, the explicit clarification of these assumptions provides a foundation on which the system can be analysed. When one knows the way in which the system is required to function, one can assess whether it is performing correctly.

5.6 Entity-Relationship Diagram for the Social Robot

Fig 7 summarises the relationships among the concepts presented in this paper, where:

- *Identity*: That which individuates robot R_i from robot R_j
- *Character*: The perceived identity of robot R_i by robot R_j
- *Stereotype*: The fundamental set of internal and external attributes that distinguish different robot types.
- *Role*: The set of subtasks robot R_i must perform in order for the group to achieve the global task
- *External Attribute*: Physical attributes of robot R_i
- *Internal Attribute*: Non-physical attributes of robot R_i

The diagram shows the complete Social Robot entity relationships that allow the development of robust complex social groups of robots capable of undertaking the global social tasks required of them. This represents the generic components of a social robot, a task, and the resulting mapping between the two. This structure allows snapshots be taken of the robot's state and the task state, therefore facilitating remote observation and analysis of the dynamic real-time system.

6. Realising the Social Robot

The realisation of a robot capable of explicit complex social behaviour or any robot, in fact, is validated by a

motivation to build a machine that can realise specific goals. It is the goal that defines the problem and in so doing defines the solutions required. The goal-oriented approach gives a criterion to assess the robot system, by seeing how effectively the robot has been at achieving its goals. The use of a behaviour-based approach results in the predefinition of the goal to be achieved. See (Brooks 1986, 1991) for the Subsumption Architecture and Duffy (2000) for explicit social behaviours in robotics. In order to achieve rational behaviour, goals are required to ascertain the success of the robot in its function, i.e. “the principle aim of a situated agent is to take action appropriate to its circumstances and goals” (Beer, 2000). When this is extended to the social domain, individual robot goals logically extend to the goals of the social group and each robot’s mutual objective in achieving the required social goal. In order to achieve this social functionality, a number of fundamental issues arise. The following elements are proposed as fundamental towards the realisation of the social robot capable of realising explicit social goals:

- Embodiment: Physical robots situated in a physical environment.
- In order for a robot to survive the dynamic complexity of real world environments, a fast reactive

process is required:

- Reflexes: Fast reactive / reflexive nature to unforeseen or unanticipated events.

As each robot is required to process both a priori knowledge and dynamic information of both physical and social environments, it must also be aware of its own internal and external attributes, and a means to store knowledge and intentionally deliberate on that knowledge:

- Deliberation: The computational machinery required to realise complex goals.
- For a robot to function as part of a social group where explicit goals are required of that group, it must have the functionality to be social:
- Social functionality: The ability of the robot to interact with other robots in its communication space.

Therefore, in order to support the development of social robots, an architecture with sufficient social and intentional functionality is required. As no research to date has dealt with the issues presented here in a coherent systematic whole, the Social Robot Architecture (Duffy, 2000; Duffy et al, 1999; Duffy et al, 1999) was developed to explicitly support strong social embodiment through the implementation of the concepts of identity,

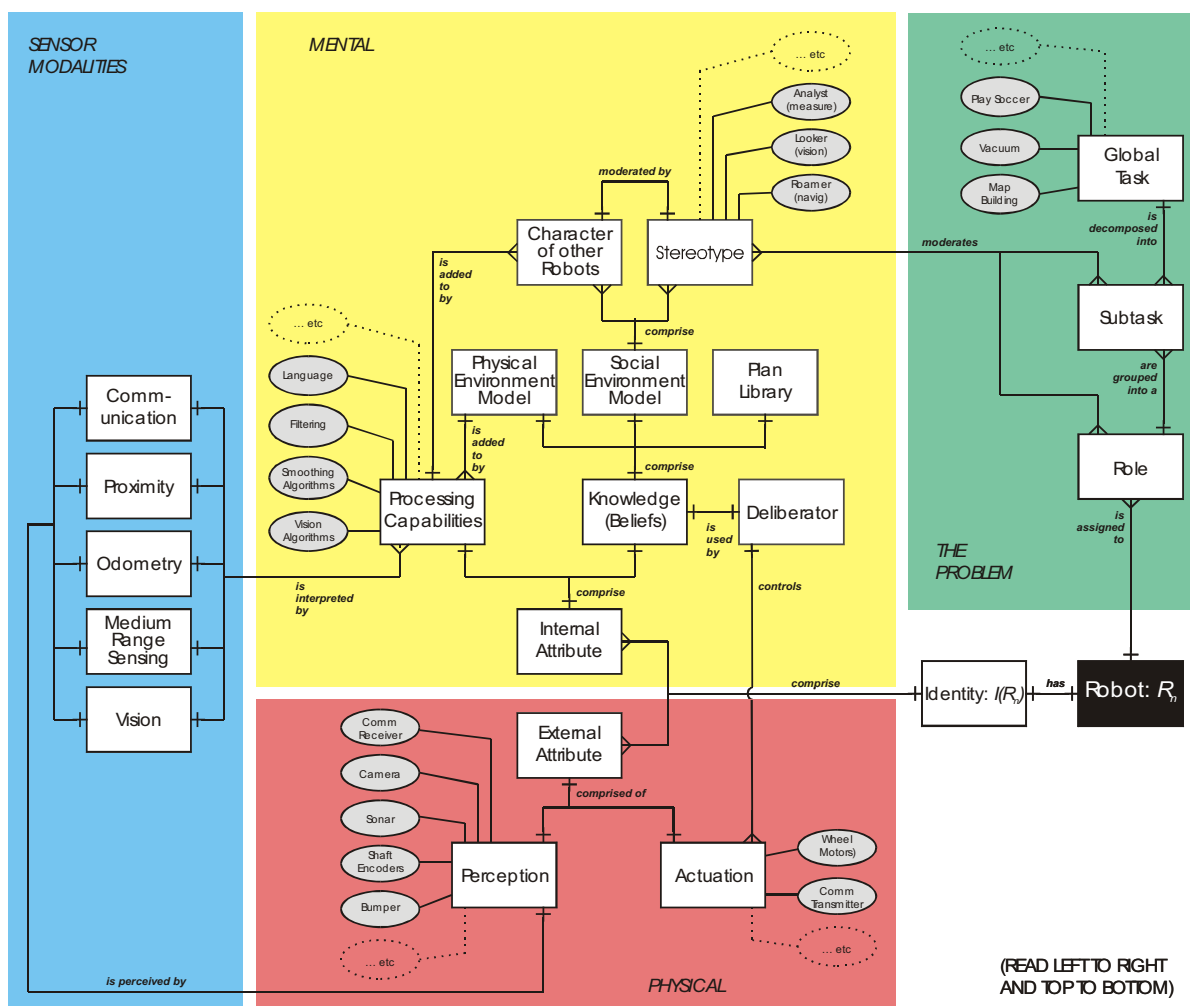


Fig. 7: The entity-relationship diagram for the Social Robot

character, stereotypes, and roles in conjunction with language and virtual reality visualisation media towards achieving truly social robots.. The architecture was developed using the agent development toolkit Agent Factory (O'Hare et al, 1998) and utilised research undertaken on the development of an agent communication language Teanga (Rooney, 2000; Rooney et al, 1999). Real-world experimentation was realised through a team of autonomous mobile Nomad Scout II robots.

The following section demonstrates how strong sociality has been realised through this framework.

7. Implementation

In order for these robots to engage in collaborative problem solving each robot must develop knowledge about its own identity and build character representations of other robots in its social environment via the use of stereotypes in order to gain a social awareness of its social environment in conjunction with its physical environment.

7.1 The Robot's Identity

Each robot has a unique identity. Its name is unique and comprises the concatenation of the robot's text name (i.e. Aengus, Bodan, Bui or Caer), the IP address of the host computer on the robot, and the time it was created:

aengus@137.43.6.4/3:01:36pm

A commitment rule is implemented whereby the robot adopts a belief about having an associated particular stereotype, e.g. scout_looker:

```
Bel(agentsName(A)) =>
Com(Self, Now, adoptBelief(stereotype(
A, scout_looker)), INACTIVE)
```

Based on this ascription, a robot can infer via deductive reasoning what internal and external attributes are associated with itself from its knowledge of stereotypes (examples of which are given in the following Section 7.2). Each robot is thus initialised at time t_0 with its name and its associated stereotype. This method is used to facilitate the modification of a robot's associated stereotype based on the addition of sensor systems for example, and the re-initialisation of its belief set of internal and external attributes corresponding to this new stereotype association. It should be noted that a robot's stereotype can thus be changed but does not evolve over time. The robot's identity evolves and develops with time, as does another robot's perception of that identity (i.e. character).

The implementation of a robot agent's identity is via Agent Factory-developed Deliberative Level of the *Social Robot Architecture* and is described in detail in Duffy (2000).

7.2 Robot Stereotypes

This section details how the robot stereotypes presented in Section 5.3 are realised. As elaborated previously, each robot has an associated stereotype and each stereotype has associated internal and external attributes.

In order that each robot can rapidly build its social model, it is equipped with a data file that acts as a "yellow pages" of the possible stereotypes that it may encounter. Each robot is thus equipped with this *a priori* knowledge about its anticipated social environment. This bootstraps the initial development of social knowledge for each robot. Should one robot encounter another with a stereotype that it does not know, it asks the robot for the information regarding the internal and external attributes of this unknown stereotype and correspondingly updates its "yellow pages". As highlighted in Section 5.5, the precondition of inherent altruism combats misinformation.

The internal and external attributes associated with each stereotype are stored as persistent beliefs within each robot's belief set, i.e. they are independent of time. Each robot can access this "yellow pages" at any time and update it accordingly. The use of commitment rules provides the procedure for robots to introduce themselves and build their character representations (see Section 7.3 for the development of character representations). This approach greatly facilitates the development of social models for any collective of socially capable robotic entities.

```
/* scout_looker_stereotype*/
/* Internal Attributes */
BelP(ia(scout_looker, social_robot_architecture))
BelP(ia(scout_looker, language_teanga))
BelP(ia(scout_looker, processing_algorithms_sonar))
BelP(ia(scout_looker, processing_algorithms_odometry))
BelP(ia(scout_looker, processing_algorithms_image))
/* External Attributes */
BelP(ea(scout_looker, nomad_scout2))
BelP(ea(scout_looker, libretto_c110_266)) [or
pc104_266, dell_laptop_266, ... etc]
BelP(ea(scout_looker, two_wheeled)) [and/or
single_arm_gripper, ... etc]
BelP(ea(scout_looker, bumper))
BelP(ea(scout_looker, sonar_ring16))
BelP(ea(scout_looker, odometry))
BelP(ea(scout_looker, camera_colour_mono))
```

Fig. 8: The ascribed internal and external attributes for a Nomad Scout II robot equipped with a vision system.

The internal and external attributes ascribed to the two stereotypes for the Nomad Scout II family of robots and used within this work are shown in figures 8 and 9. The first corresponds to a Nomad Scout II robot equipped with a vision system (Duffy, 2000) and corresponds to the following fundamental internal and external features:

- Internal: - Deliberative model (i.e. Social Robot Architecture)
 - Communication language,
- Perception processing algorithms: sonar, odometry, vision
- External: - Robot type¹

¹ This pertains to the physical construction information of the robot, i.e. its dimensions and weight.

- Processing capabilities
- Actuators: wheels
- Perceptors: bumper, sonar, odometry, vision

This information is stored in each robot's belief set as shown in fig. 8.

As a number of the Nomad Scouts used in this work are not currently equipped with vision systems, those that do have the associated `scout_worker` stereotype (see fig. 9). This methodology extends to any autonomous mobile robot equipped with the Social Robot Architecture and the necessary communication hardware (through wireless communication capability). Each robot has an ascribed stereotype which it communicates on "meeting" other robots in its social envelope. The issue of social embodiment is therefore explicitly addressed. As the architecture is clearly designed to be hardware independent, the only requirements of a suitable hardware platform is adequate processing capabilities in the form of a PC, Apple Macintosh, or Unix-based processor. The corresponding stereotypes would subsume the functionality available on such platforms.

```

/* scout_worker_stereotype*/
/* Internal Attributes */
BelP(ia(scout_worker,social_robot_architectur
e))
BelP(ia(scout_worker,language_teanga))
BelP(ia(scout_worker,processing_algorithms_so
nar))
BelP(ia(scout_worker,processing_algorithms_od
ometry))
/* External Attributes */
BelP(ea(scout_worker,nomad_scout2))
BelP(ea(scout_worker,libretto_c110_266))[or
pc104_266, dell_laptop_266, ... etc]
BelP(ea(scout_worker,two_wheeled)) [and/or
single_arm_gripper, ... etc]
BelP(ea(scout_worker,bumper))
BelP(ea(scout_worker,sonar_ring16))
BelP(ea(scout_worker,odometry))

```

Fig. 9. The ascribed internal and external attributes for a basic Nomad Scout II robot.

The general introduction of additional hardware or software components to a class of robots like the Nomad Scout II simply involves the addition of the corresponding internal or external attribute to the robots stereotype. The use of stereotypes both facilitates the social model building and provides a medium through which task allocation can be achieved. As the structure is in essence relatively simple, its inherent robustness provides a considerable degree of reliability in the development and maintenance of social relationships. The process whereby the Deliberative Level of the Social Robot Architecture (Duffy *et al*, 1999; Duffy, 2000) develops these social relationships is through the use of commitment rules. These rules provide the deliberative machinery for the development of the social models.

7.3 Commitment Rules

The Commitment Rules are a fundamental part of the Deliberative Level that is developed by Agent Factory and constitute the rules through which the robot will

commit to a certain action based on active beliefs in its belief set (Duffy, 2000). Examples of such are `adoptBelief`, `dropBelief`, `inform`, `request`, `commit`.

As mentioned previously, each robot must introduce itself to any other robot in the group on joining a group. When a robot communicates its associated *stereotype*, any other robot must associate the appropriate internal and external attributes to that robot based on the attributes ascribed to this stereotype. The following commitment rules represent the ascription of appropriate beliefs to each robots stereotype, even the robot SELF:

```

Bel(stereotype(A,S)) & Bel(ea(S,E))
=> Com(Self,Now,adoptBelief(
  Bel(ea(E,A)), INACTIVE)

```

```

Bel(stereotype(A,S)) & Bel(ia(S,I))
=> Com(Self,Now,adoptBelief(
  Bel(ia(I,A)), INACTIVE)

```

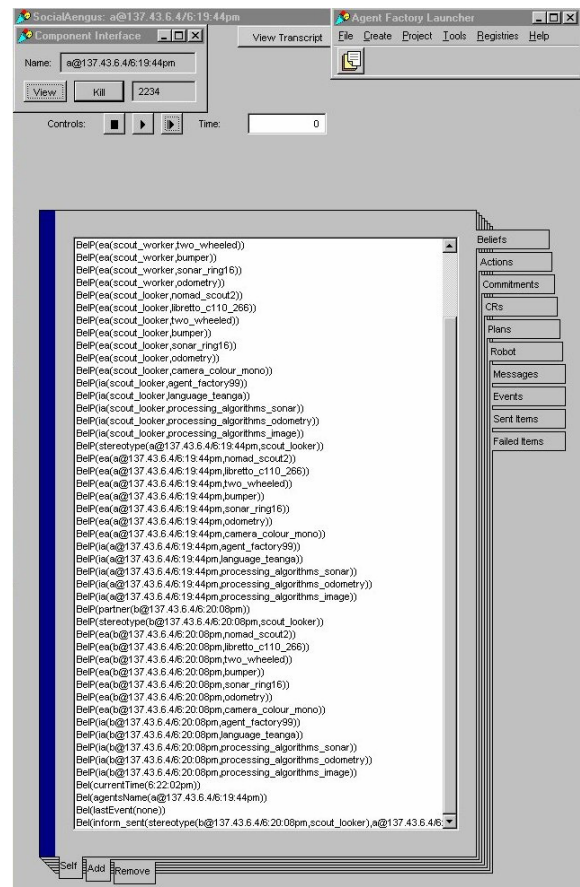


Fig. 10. The belief set of RobotA showing its social model (*character* representation - bootstrapped by *stereotype* information) of RobotB

In order for one robot to develop an internal representation of other robots, it requires each robot's stereotype. Conversely it must communicate its own stereotype to others:

```

Bel(partner(A)) & Bel(agentsName(B))
& Bel(stereotype(B,S)) =>
Com(Self,Now,inform(B,stereotype(
  A,S)), INACTIVE)

```


Each robot can interpret the stereotype it receives from other robots, associate it with that robot's name. Then based on the commitment rules given above and each robot's knowledge of the internal and external attributes ascribed to each stereotype, a robot can develop beliefs about another's internal and external attributes:

```
Bel(informed_belief(stereotype(A,S),
  B)) & Bel(partner(B)) =>
  Com(Self,Now,adoptBelief(stereotype(A,S),INACTIVE))
```

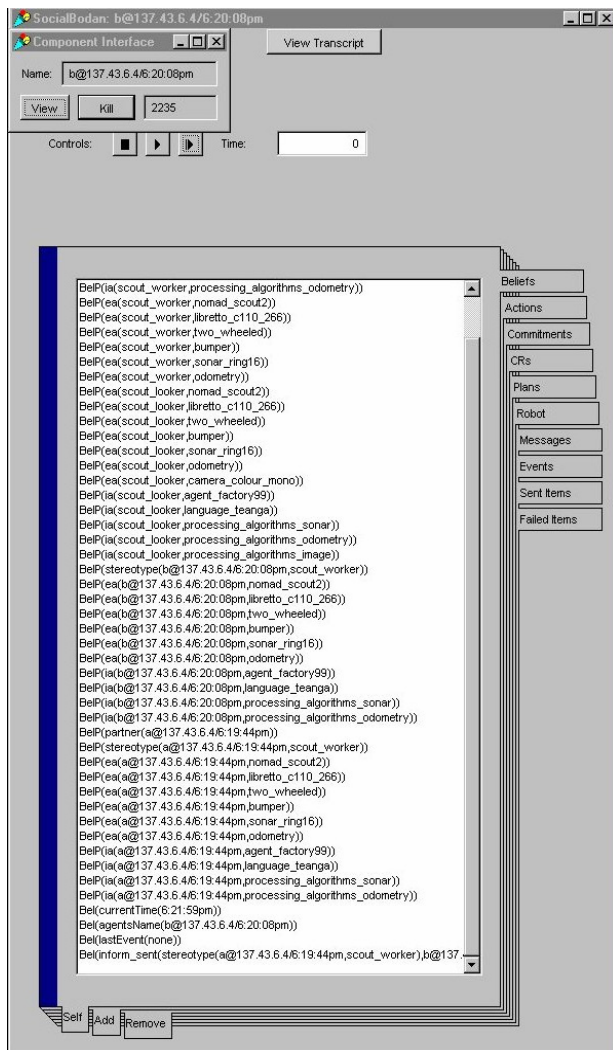


Fig. 11. The belief set of RobotB showing its social model (*character* representation - bootstrapped by *stereotype* information) of RobotA

Thus each robot has a concept of self, i.e. its identity, as well as representations of the other robots in its social environment.

Figures 10 and 11 show the generated social model for two socially communicating robots, RobotA and RobotB. Each robot has developed a notion of *identity* based on its knowledge of its own internal and external attributes and also a *character* representation for the other robot (via *stereotype* knowledge).

Based on each robot's knowledge about the attributes for particular stereotypes, each robot adopts beliefs about its own attributes by knowing its ascribed stereotype and also builds representations for all other robots in its social environment. Its identity therefore constitutes its complete knowledge about both itself and its social and physical environment. As a robot's knowledge of its environment (both physical and social) develops over time through interaction and communication, each robot's identity also develops and evolves. This evolution is firmly grounded on its associated stereotype in order to provide a consistent benchmark for both itself and how other robots develop character representations of its inherent internal and external attributes.

8. Discussion

The previous sections have addressed the basic issues presented in Section 5, that of embracing both physical and social embodiment within a collective of autonomous mobile robots undertaking explicit social tasks. The conceptual notions discussed in Sections 5 and 6 in the form of a socially empowered autonomous mobile robot, the *Social Robot*, have been realised (Duffy, 2000; Duffy *et al*, 2000]. The functionality of the *Social Robot Architecture* and its constituent parts demonstrate the ease with which socially empowered robotic entities can be realised. Each robot has an *identity*, an associated *stereotype* based on its internal and external attributes and the capacity to build *character* representations of other robots in its social environment. A degree of social embodiment is thus achieved. Each robot can therefore collaborate and coordinate their efforts in performing a complex social task through social representations of its colleagues and the Agent Communication Language *Teanga*. Experimentation has demonstrated the power of this approach in both untailored office environments and tailored soccer playing environments in achieving coherent social interaction in real-world scenarios via a team of four Nomad Scout II robots.

These experiments require that a number of cooperating robots explicitly collaborate to perform each social task. While there is the question as to whether many robots can complete a task more efficiently than a single robot, this issue is not addressed in this work. The benefits of a single robot over multiple robots are highly dependent on the tasks undertaken. The objective was rather to implement the *Social Robot Architecture*, and more generally to realise the notion of a Social Robot, by addressing the issues of both physical and social embodiment. As such, the tasks that the robots are required to address are specifically oriented to multiple robotic entities working in mutual cooperation. The realisation of the social tasks required of the collective of robots is not based on emergent social behaviours as found in (Mitsumoto *et al*, 1995; Cai *et al*, 1995; Cao *et al*, 1997), but rather on explicit coordination and cooperation.

These experimental tasks demonstrate how the social features presented in section 5 and illustrated in section 7 were utilised for social complex tasks undertaken by a team of Social Robots.

9. Conclusion

Physical embodiment is a fundamental component of intelligence. When a physical robot is situated in an environment occupied by other physical robots, social embodiment exists. Physical embodiment and social embodiment are inherently linked. While generally construed as extremely complex, the embodiment of such complex notions as social functionality in conjunction with AI-based robot control strategies on physical autonomous mobile robots may provide us with an alternative understanding of the fundamentals of an artificially intelligent entity. In returning to the issues raised in Section 2, being “social” implies the existence of interactive relationships. In order for multiple robots to exhibit and maintain robust behaviours in a mutual environment, a degree of social functionality is necessary. A robot with the capacity to interact socially with other robots in this social environment has an obvious advantage of being better equipped to coordinate and collaborate its behaviour with others and in so doing reinforces its survival in complex unpredictable real world scenarios. The tools required to achieve such social interaction have been presented in this work. A “Social Robot” has been realised.

A new issue arises when robots are required to function as part of a social community. As language, in this case the ACL Teánga (Rooney, 2000), provides the medium for social communication, the argument for a more “natural” language-based cognitive process is presented, similar to that found in BDI-based approaches [Rao & Georgeff, 1991; Duffy, 1999; Duffy, 2000]. Sensory information must be interpreted and understood at the deliberative level in explicit control methodologies. This enforces a degree of abstraction from sensor voltages and pixel maps of vision systems and the generation of such abstract notions as *agent_events*, (for example, *wall_from(x₁,y₁,x₂,y₂,sonar_number,distance)*). Similarly the actuator functions adopt a comparable format, e.g. *move_forward(x₁,y₁)*. When the deliberative level in such an architecture allows cognitive functions be processed in a more natural language format, the sentence structure communicated to other robots in its social environment becomes more expressive. Obviously a degree of ambiguity may arise and must be addressed [Rooney, 2000]. The existence of a robot in a social community therefore revalidates using such deliberative engines as found in BDI approaches and supports the notions of beliefs, desires and intentions in a social environment.

Similarly, when such a system’s social and environmental knowledge is stored in the form of beliefs, a degree of system observability is obtained. Limited decoding of this knowledge is required for interpretation by any human observer. This facilitates an understanding

of the robot’s data acquisition and learning processes for testing and analysis. The social and physical representations that each robot builds of its environment become more transparent. An interesting future work will involve the development of a speech recognition system and voice synthesiser to enable direct social communication with people. The intentional architecture could greatly facilitate the human-robot communication process when it already functions at an intentional level through its BDI functionality.

The Social Robot framework therefore constitutes a next stage in the evolution of explicit autonomous mobile robot control by extending research in the multi-robot task domain as well as providing a generic approach for physically and socially embodied robots.

10. References

- Baumeister, R. (1987). How the self became a problem: A psychological review of historical research. *Journal of Personality and Social Psychology* 52, 163-176.
- Beer, R.D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences* 4(3), 91-99.
- Brooks, R.A. (1986). A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automaton* 2(1).
- Brooks, R.A. (1991). Intelligence Without Representation. *Artificial Intelligence Journal* 47, 139-159.
- Cai, A, Fukuda, T., Arai, F., Ueyama, T. & Sakai, A. (1995). Hierarchical Control Architecture for Cellular Robotic System – Simulation and Experiments. *IEEE International Conference On Robotics and Automation*.
- Cao, Y.U., Fukunaga, A.S., Kahng, A.B. (1997). Cooperative Mobile Robotics: Antecedents and Directions. *Autonomous Robots* 4, 1-23.
- Castelfranchi, C. (1990). Social power. A point missed in multi agent, DAI and HCI. In: Demazeau, Y. & Mueller, J.P. (Eds.), *Decentralized AI*. Amsterdam: Elsevier.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. MIT Press.
- Cohen, P.R., Levesque, H.J. (1997). On team formation. In: Hintikka, J. & Tuomela, R. (Eds.), *Contemporary Action Theory*. Synthese.
- Conte, R., Miceli, M., & Castelfranchi, C. (1991). Limits and levels of cooperation: Disentangling various forms of social interaction. In: Demazeau, Y. & Mueller, J.P. (Eds.), *Decentralized AI – 2*. Amsterdam: Elsevier.
- Donald, B.R. (1994). Analysing teams of cooperating mobile robots. Internal Report TR94-1429, Dept. of Computer Science, Cornell University, June.
- Duffy, B.R. (2000). *The Social Robot*. PhD Thesis, Department of Computer Science, University College Dublin, November.

- Duffy, B.R. & Joue, G. (2000). Intelligent Robots: The Question of Embodiment. Proceedings of BRAIN-MACHINE'2000, 20-22 December, Ankara, Turkey.
- Duffy, B.R., O'Hare, G.M.P., O'Donoghue, R.P.S., Rooney, C.F.B. & Collier, R.W. (1999). Reality and virtual reality in mobile robotics. Proceedings of the 1st International Workshop on Managing Interactions in Smart Environments MANSE'99, Dublin, December.
- Duffy, B.R., Collier, R.W., O'Hare, G. M. P., Rooney, C.F.B. & O'Donoghue, R.P.S. (1999). SOCIAL ROBOTICS: Reality and Virtuality in Agent-Based Robotics. Proceedings of Bar-Ilan Symposium on the Foundations of Artificial Intelligence: Bridging Theory and Practice (BISFAI99).
- Durfee, E.H., Lesser, V.R. (1987). Using partial global plans to coordinate distributed problem solvers. Proceedings of the Tenth International Joint Conference on Artificial Intelligence, August.
- Fukuda, T., Nakagawa, S., Kawauchi, Y. & Buss, M. (1989). Structure Decision for Self Organising Robots Based on Cell Structures. IEEE: Robotics & Automation, Scottsdale Arizona.
- Grosz, B.J. (1996). Collaborative systems. *AI Magazine* 17(2), 67-85.
- Haddadi, A. (1995). Reasoning about cooperation in agent systems: A pragmatic Theory. Ph.D. Thesis, University of Manchester Institute of Science and Technology (UMIST), March.
- Holland, O.E., Deneubourg, J.-L. (2000). From local actions to global tasks: Stigmergy and collective robotics. In: Ritter, H., Cruse, H. & Dean, J. (Eds.), *Prerational intelligence: Adaptive behavior and intelligent systems without symbols and logic* (Vol. 2, pp.549-563). Kluwer Academic Publishers.
- Hoyle, R. (1989). Self and Identity in Everyday Life. International Society for Self and Identity, <http://www.soton.ac.uk/~gramzow/ISSI/daily.htm>
- Jennings, N.R. (1993). Specification and Implementation of a Belief- Desire-Joint-Intention Architecture for Collaborative Problem Solving. *International Journal of Intelligent and Cooperative Information Systems* 2(3), 289-318.
- Jennings, N.R. & Campos J.R. (1997). Towards a Social Level Characterisation of Socially Responsible Agents. *IEEE Proceedings on Software Engineering* 144(1), 11-25.
- Kinny, D., Georgeff, M. & Rao, A. (1996). A methodology and modelling technique for systems of BDI agents. Proceedings of 7th European Workshop on Modelling Autonomous Agents in Multi-Agent Worlds (MAAMAW'96), LNAI 1038 (pp.56-71). Springer-Verlag.
- Lesser, V., Decker, K., Carver, N., Garvey, A., Neimen, D., Prasad, M.V. & Wagner, T. (1998). Evolution of the GPGP Domain Independent Coordination Framework. (A version is available as UMASS CS TR #1998-005).
- Mitsumoto, N., Hattori, T., Idogaki, T., Fukuda, T. & Arai, F. (1995). Self-organising micro robotic system. Proceedings of the 6th International symposium on Micro Machine and Human Science. IEEE.
- Newell, A. (1990). Unified theories of cognition. Harvard University Press.
- Newell, A. (1982). The Knowledge Level. *Artificial Intelligence* 18, 87-127.
- O'Hare, G.M.P., Collier, R., Conlon, J. & Abbas, S. (1998). Agent Factory: An Environment for Constructing and Visualising Agent Communities. Proceedings of AICS98.
- O'Hare, G.M.P., Jennings, N.R. (Eds.). (1996). *Foundations of Distributed Artificial Intelligence*. New York: Wiley Interscience Publishers.
- Panzarasa, P., Norman, T.J. & Jennings, N.R. (1999). Modeling sociality in the BDI framework, submitted to *World Scientific*.
- Rao, A.S. & Georgeff, M.P. (1989). Modelling Rational Agents within a BDI Architecture. *Principles of Knowledge Representation & Reasoning*. San Mateo, CA.
- Rooney, C. (2000). A formal semantics for Teánga. Technical Report No. UCD-PRISM-00-05, Department of Computer Science, University College Dublin.
- Rooney, C.F.B., Collier, R.W., O'Hare, G.M.P., Duffy, B.R., O'Donoghue, R.P.S., (1999) "Teánga: A Language for Social Robots" *Cognitive Science For The New Millenium*, Dublin.
- Sekiyama, K. & Fukuda, T. (1999). Towards social robotics. *Applied Artificial Intelligence* 13, 213-238.
- Sharkey, N. & Zeimke, T. (1998). A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science* 10, 361-391.
- Shoham, Y. (1993). Agent-oriented programming. *Artificial Intelligence* 60(1), 51-92.
- Smithers, T. (1995). On quantitative performance measures of robot behaviour. *Journal of Robotics and Autonomous Systems* (Special Issue: The Biology & Technology of Intelligent Autonomous Agents) 15, 107-133.
- Stone, P. & Veloso, M. (1999). Task Decomposition and Dynamic Role Assignment for Real-Time Strategic Teamwork, *Intelligent Agents V*. In Mueller, Singh, and Rao (Eds.), *Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL'98)*. Heidelberg: Springer Verlag

Acknowledgements

The support of Enterprise Ireland through grant No.'s SC/96/621 COMEDY and SP/97/08 IMPACT (Initiative for Mobile Programmable Agent-based Computing Technologies) is gratefully acknowledged.