Handling robot sociality: a goal-based normative approach

Patrizia Ribino, Carmelo Lodato, and Ignazio Infantino

Istituto di CAlcolo e Reti ad Alte prestazioni (ICAR)
Consiglio Nazionale delle Ricerche, Italy
{patrizia.ribino,carmelo.lodato,ignazio.infantino}@icar.cnr.it

Abstract. The increasing development of service robots devoted to various functions arisen the need to demonstrate additional social capabilities beyond their primary functionality. For improving robot sociality, among other abilities, robots need to implement the capability to interact with humans using the same principles as humans do following social norms. In this work, we propose an extension of a goal-based normative framework to cover new abstractions such as qualitative goals, social norms, and expectations which constitute essential elements for handling robot sociality. Moreover, we have integrated such extended normative framework into a Nao robot platform. An implementation of the proposed framework is described and tested in a simulated environment.

Keywords: social robots, social norms, normative reasoning

1 Introduction

As result of the increasing progress in the field of the Artificial Intelligence, robots are expected to become more and more available in everyday environments. Among several issues, the integration of robots into the society depends on their capability to demonstrate socially acceptable behaviours to be perceived by humans as suitable partners in collaborations. To define socially acceptable actions, we refer to the branch of socio-cognitive theory that has documented the existence of two orthogonal dimensions in social judgement [1, 2].

The social judgement of an individual can be represented through two components: *social utility* and *social desirability*. Social utility refers to individuals' capacity to satisfy the functional requirements of a given social environment. It varies along an incompetence-to-competence horizontal axis that corresponds to the perceived ability of the social target to reach social success. It pertains to adaptive traits like skilled/unskilled, proactive/passive. For example, self-sufficiency [2] and being focused on goal achievement [3] are perceived as socially useful behaviours. On the other hand, social desirability refers to the degree of likeableness of a person in his/her relationships with others in a given social environment. It varies along an unlikability-to-likability vertical axis, that corresponds to the perceived ability of the social target to gain social approval. It concerns aspects such as polite, honesty, respect, only to cite a few.

According to Sommet *et.al* [4], socially useful behaviours are typically those described as focused on the self for sustaining the practical fulfilment of one's goals. Conversely, socially desirable behaviours are those defined as directed on the others, that involve benevolent interaction styles. From humans perspective, showing the social utility of a robot can be easier than perceiving him as socially desirable because in this latter case it is necessary to demonstrate additional social capabilities beyond robot's primary functionality. Indeed, the social utility of service robots deployed for various functions in public spaces such as airports, hospitals, logistic warehouses is readily perceived by humans. Conversely, to be socially desirable, robots [5] need to show not only "human social" features (like the expression of emotions, ability to conduct high-level dialogue, to develop personality and social competencies and so on), but also capabilities to interact using the same principles as humans do. As it arisen from cognitive and social science, human interactions are fundamentally based on normative principles. For example, many forms of interaction are institutionalised and pertain to the political and economic structures of the society that are defined by rules and prescribed by laws that enforce behaviour [6]. Other types of interaction are based on conventions, such as on what side of the road people should drive [7]. Finally, most human interactions are often influenced by more profound social and cultural standards, so-called social norms [8]. A social norm is commonly seen as [8]: *a rule of behaviour such that individuals prefer to conform to it on condition that they believe that (a) most people in their relevant network conform to it, and (b) most people in their relevant network believe they ought to conform to it and may sanction deviations.* Norms directly identify possible actions as desirable or undesirable in a given community and a particular context involving social expectations and guiding the choice of people' actions [9, 8]. Social expectations are people's beliefs about other people's behaviours and beliefs in certain situations. Beyond social norms, expectations play an important role in regulating social behaviours. Indeed, an individual may comply with social norms in the presence of relevant expectations, but (s)he does not follow them in their absence [8].

A current challenge is how to incorporate norm processing into robotic architectures, because it requires addressing several issues such as the specification of social norms, how they can be activated, how social plans can be generated for expressing social behaviours, the conflicts resolution and the acquisition of new norms [9]. The contribution of this paper is to address the social desirability as the ability of a robot to show itself as conforming to social norms. In so doing, we extended the approach we have presented in [10] by introducing the concepts of qualitative goals and revising the concept of achievement goals for modelling different objectives of social robots. Then, we extended the definition of norms to cover the peculiarities of social norms by introducing desirable operators and expectations. Finally, we have integrated our normative framework into a robotic platform. An implementation of our approach is tested in a simulated humanoid robot Nao by using Choreographe [11], a user-friendly application for controlling robots, creating behaviours and accessing data acquired by the sensors.

The rest of the paper is organised as follows. Section 2 presents an overview of the related works. Section 3 presents the theoretical foundations of the proposed approach. Then, in Section 4 a case study about robot sociality is presented. Finally, in Section 5 conclusions are drawn.

2 Related Works

Shortly, social robots will play an ever more significant role, working for and in cooperation with humans. In so doing, they should show social capabilities [5] such as interacting with humans naturally. An emerging challenge is to provide a robot with the normative reasoning to behave in compliance with the same social norms as humans do. To the best of our knowledge, only a few recent works address such issue in an explicit and general way, and a lot of work must still be done to incorporate sophisticated norm processing into robotic architecture. In [12], authors present a framework for planning and execution of social plans, in which social norms are explicitly represented in a domain and language-independent form. In [13], Brinck *et.al* discuss the role of the social norms in the design of human-robot interactions focusing on the dynamic information that a robot needs to comply with social norms. In particular, they pay attention to three elements (gaze and face, place in space, and orientation, posture and movement) that are important as sources of social information. An initial step toward a cognitive-computational model of norms by delineating core properties of the human norm system, contrasting two models of a computational norm system, and deriving implications for how robotic architectures would implement such a norm system is described in [9].In such work, authors focus on modelling norms as directly and indirectly connected networks discussing mechanisms of co-activation of rules that are connected to other norms. Finally, in [14] an approach to creating a computational model of social norms based on identifying some values that are considered relevant in some culture. Appropriate metrics quantify such values. Social norms are used as the requirement for maximising such metrics. In so doing, authors introduce a model for concrete beliefs of the actors that are relevant to the social scene.

In this work, we propose a normative approach that allows exploiting the advantages of goal modelling to make social robots able to reason about dynamic situations pro-actively. In so doing, we suggest the concept of quality goals for modelling the pursuit of social values by a robot. Then we define social norms by introducing desirability operators for representing preferences about acceptable behaviours. Finally, we define the expectations formally as a new mental concept a robot sees as a motivator for pursuing social values by following social norms.

3 Goal-based normative framework for social robots

A widely-accepted approach for developing intelligent agents (both robots and bodiless agents) is a cognitive approach, where agents are modelled using mental concepts such as beliefs, goals, plans, rules and so on. Among them, to develop agents able to reason about dynamic contexts pro-actively, a fundamental abstraction is the concept of *goal*. Rich literature addresses issues about goal modelling for intelligent agents, defining a great variety of goal's types [15]. The *achievement goal* is the most used kind of goal. It models the most recurring functional requirements of this kind of systems. The well known cognitive definition is [15]: an *achievement goal* represents a desired state of the world that an agent wants to reach.

On the other side, for providing intelligent agents with normative reasoning, a fundamental abstraction to be modelled is the concept of the social norm. Social norms are behavioural rules considered acceptable in determined contexts, which refer to the standard of *desirability* in a community. Thus, social norms are behavioural expressions of *abstract social values* (such as politeness, dignity, hospitality, honesty, etc.) that underlie the preferences in a group in various situations. For example, the norm "*everyone should to queue to ticket office*" involves the social values of equality, efficiency and respect for orderliness. In other words, widely accepted social values provide the grounds for complying or rejecting certain behavioural norms.

Among several functions that social norms serve in the society, they mainly provide guidelines for *expected* modes of social behaviour. Thus, for keeping society functioning, an important role is played not only by the direct rules but also by the *expectations* about the conduct of the members of the society. If few members of the group follow the norm (e.g., do not use a cellphone during class), then the norm is weakened, and it may be no longer treated as binding. If few members of the group expect others to follow the instruction, it becomes optional and loses its character as a norm. These peculiarities distinguish norms from goals because the latter can hold even when individuals disregard entirely other community expectations.

A cognitive definition of social norm [16, 9] states that:

- An agent represents the instruction to [not] perform a specific action or general class of action.
- An agent believes that a number of individuals in the group in fact (do not) follow the norm.
- An agent believes that a sufficient number of individuals in the groups expects others in the group to (not) follow the norm.

In this work, we want to add a further condition to the previous ones. So that a social robot conforms its behaviour to social norms of a group as humans do, it has also to share the same social values as the members of that community. Thus, we extend the previous definition with the following condition:

- An agent wants to pursue a social value.

To incorporate norm processing into social robots, we propose a goal-based normative approach that extends our previous work [10] for covering different aspects both functional and non-functional of a social robot. In particular, to implement the feature of sociality, we extended the definition of the norm to cover also the peculiarities of the *social norms* by introducing the concept of *expectation*. Moreover, we have proposed the notion of qualitative goals for modelling the pursuit of social values.

Before defining norms and goals, we need to introduce the definition of the state of the world [10] that is fundamental for the following. The *state of the world* represents a set of declarative information about events occurring within the environment and relations among events at a specific time. An event can be defined as the occurrence of some fact that can be perceived by or be communicated to an intelligent agent. Events can be used to represent any information that can characterise the situation of an interacting user as well as a set of circumstances in which the intelligent agent operates at a specific time.

Definition 1 (State of the world).
Let $\mathcal{D}$ be the set of concepts defining a domain. Let $\mathcal{L}$ be a first-order logic defined on $\mathcal{D}$ with $\top$ a tautology and $\bot$ a logical contradiction, where an atomic formula $p(t_1, t_2..., t_n) \in \mathcal{L}$ is represented by a predicate applied to a tuple of terms $(t_1, t_2..., t_n) \in \mathcal{D}$ and the predicate is a property of or relation between such terms that can be true or false.

A *state of the world* in a given time t ($\mathcal{W}^t$) is a subset of atomic formulae whose values are true at the time t:

$$\mathcal{W}^t = [p_1(t_1, t_2, ..., t_h), ..., p_n(t_1, t_2, ..., t_m)]$$

Definition 1 is based on the close world hypothesis that assumes all facts that are not in the state of the world are considered false. In the next sections, we introduce the elements of the proposed approach.

3.1 Types of Goals

An *Achievement goal* represents the desired state that has to be achieved. They express goals which are not currently fulfilled, and which the agent, pursuing the appropriate actions, acts to reach them. To define, an achievement goal we extended the general definition of goal proposed in [10].

Definition 2 (Achievement Goal). Let $\mathcal{D}$, $\mathcal{L}$ and $p(t_1, t_2..., t_n) \in \mathcal{L}$ be as previously introduced in the definition 1. Let $t_c \in \mathcal{L}$, $f_s \in \mathcal{L}$ and $f_c \in \mathcal{L}$ be formulae that may be composed of atomic formulae by means of logic connectives AND($\wedge$), OR ($\vee$) and NOT ($\neg$). An *Achievement Goal* is a triple $\langle t_c, f_s, f_c \rangle$ where t_c (*trigger condition*) is a condition to evaluate over a state of the world $\mathcal{W}^t$ when the goal may be actively pursued, f_s (*final state*) is a condition to evaluate over a state of the world $W^{t+\Delta t}$ when it is eventually addressed, f_c (*failure condition*) is a condition to evaluate over a state of the world $W^{t+\Delta t}$ when the goal is no longer applicable. An achievement goal is:

i) active if $t_c(\mathcal{W}^t) \wedge \neg f_s(\mathcal{W}^t) = true$
ii) addressed if $f_s(\mathcal{W}^{t+\Delta t}) = true$
iii) is dropped if $f_c(\mathcal{W}^{t+\Delta t}) = true$

On the contrary, a qualitative goal is a kind of goal that is perceived more than fulfilled. It is a goal for which satisfaction criteria are not defined in a clear-cut way.

Definition 3 (Qualitative goal). Let $\mathcal{D}$, $\mathcal{L}$ and $p(t_1, t_2..., t_n) \in \mathcal{L}$ be as previously introduced in the definition 1. Let $t_c \in \mathcal{L}$ and $f_s \in \mathcal{L}$ be formulae that may be composed of atomic formulae by means of logic connectives AND($\wedge$), OR ($\vee$) and NOT ($\neg$).

An *qualitative goal* is a tuple $\langle t_c, q_s, s_c, f_c, \rangle$ where t_c (*trigger condition*) is a condition to evaluate over a state of the world $\mathcal{W}^t$ when the quality goal may be actively pursued, q_s (*qualitative state*) is the state to head toward, s_c(*suspending condition*) is a condition to evaluate over a state of the world $W^{t+\Delta t}$ when the quality goal has to be suspended, f_c (*failure condition*) is the condition to evaluate over a state of the world $W^{t+\Delta t}$ when the goal is no longer applicable. A qualitative goal is:

i) active if $t_c(\mathcal{W}^t) \wedge \neg m_c(\mathcal{W}^t) \wedge \neg s_c(\mathcal{W}^t) \wedge \neg f_c(\mathcal{W}^t) = true$
ii) suspended if $s_c(\mathcal{W}^{t+\Delta t}) \wedge \neg f_c(\mathcal{W}^{t+\Delta t}) = true$
iii) dropped if $f_c(\mathcal{W}^{t+\Delta t}) = true$

A qualitative goal never quite reach the state it is heading toward, but instead get closer and closer. Thus, a qualitative goal when activated will be continuously pursued until it is suspended or dropped.

In the context of social robots, the concept of achievement goals allows us to represent functional requirements that a social robot has to be able to satisfy. Conversely, a qualitative goal enables us to describe the pursuit of a social value that can not be described by mean a clear condition to be reached. An agent has to continuously perform actions that give positive contributions to sustaining a quality state.

3.2 Social Norms

In the following, we provide an explicit representation of social norms and robot's expectations. In particular, we adapt the definition of norm presented in [10] for representing social norms introducing the desirability operators: *Desirable*, *Undesirable* and *Indifferent* [17]. Desirability operators represent preference in a wide sense. The following principles underpin the desirable operators:

$$\rho \leftrightarrow \psi \quad then \quad Des\ \rho \leftrightarrow Des\ \psi \tag{1}$$

$$Des\ \rho \rightarrow \neg Des\ \neg\rho \tag{2}$$

$$Des\ \rho \rightarrow \neg Undes\ \rho \tag{3}$$

In our context, ρ designates a proposition that asserts that an act or a state of affair is done or reached. Thus, $Des\ \rho$ is read as "it is desirable that the situation described by the descriptive sentence ρ is realised". In particular, (3) expresses that something cannot be desirable and undesirable at the same time.

Definition 4 (Social Norm). Let $\mathcal{D}$, $\mathcal{L}$, and $p(t_1, t_2..., t_n) \in \mathcal{L}$ be as previously introduced in the definition 1. Let $\phi \in \mathcal{L}$ and $\rho \in \mathcal{L}$ be formulae that may be composed of atomic formula by means of logic connectives AND($\wedge$), OR ($\vee$) and NOT ($\neg$). Moreover, let $Des_{op} = \{Desirable, Undesirable, Indifferent\}$ be the set of desirability operators. A *Social Norm* is defined by the elements of the following tuple:

$$n = \langle p, qs, \rho, \phi, d \rangle \quad \textit{where}$$

- p is a *Position* the norm refers to. A Position indicates a status of an individual in a society. The symbol _ means that norms refers to anyone.
- qs is a *quality state*. It represents the social value the norm underlying.
- $\rho \in \mathcal{L}$ is a formula expressing the set of actions and/or state of affairs that the norm disciplines.
- $\phi \in \mathcal{L}$ is a logic condition (to evaluate over a state of the world $\mathcal{W}^t$) under which the norm is applicable.
- $d \in D_{op}$ is the desirability operator applied to ρ that the norm prescribes for sustaining the quality state qs in a state of the world $\mathcal{W}^{t+\Delta t}$:

$$d(\rho) \equiv \begin{cases} \rho(\mathcal{W}^{t+\Delta t}) = true & \text{if } d = Desirable \\ \neg\rho(\mathcal{W}^{t+\Delta t}) = true & \text{if } d = Undesirable \\ \rho(\mathcal{W}^{t+\Delta t}) \vee \neg\rho(\mathcal{W}^{t+\Delta t}) = true & \text{if } d = Indifferent \end{cases} \tag{4}$$

Let us consider a society where the politeness is considered a shared social value. Thus, a social norm such as the following "*It is desirable that a guy gives up his seat if an elderly person is standing up*" prescribes an acceptable behaviour. According to the definition 5, the previous norm is applied to a guy (i.e., the position) and it prescribes that, in a given state of the world where an elderly person is standing up ($i.e. : \phi(\mathcal{W}^t) = true$) then the action "give up own seat" is expected to be true in a consecutive state of the world ($i.e. : \rho(\mathcal{W}^{t+\Delta t}) = true$).

As said before, a further important role is played by the expectations. By definition, the preference to conform to a social norm is conditional, it implies that one may comply with a social norm in the presence of the relevant expectations, but do not obey the norm in the absence of such expectations. We initially considered such expectations in a broad sense as a motivator for pursuing a social value, thus leading an agent to follow the related social norms. Conversely, repeated negative feedbacks about its expectations cause the loss interest in that social value, thus ignoring social norms [1]. In this work, an expectation is generated by certain circumstances, and it is satisfied when the expected state is true in a consecutive state of the world before its time to fulfil (if any).

DEFINITION 5 (EXPECTATIONS). Let $\mathcal{D}$, $\mathcal{L}$, and $p(t_1, t_2..., t_n) \in \mathcal{L}$ be as previously introduced in the definition 1. Let $n = \langle p, qs, \rho, \phi, d \rangle$ be a *Social Norm*. Let $t_c \in \mathcal{L}$ and $e_s \in \mathcal{L}$ be formulae that may be composed of atomic formulae by means of logic connectives AND($\wedge$), OR ($\vee$) and NOT ($\neg$). An *Expectation* is a couple $\langle n, e_s \rangle$ where $n = \langle p, qs, \rho, \phi, d \rangle$ is the social norm generating the expectation and e_s (*expected state*) is a condition to evaluate over a state of the world $W^{t+\Delta t}$ when expectation is eventually satisfied. Moreover, an expectation may have time to fulfilled (ttf), that is the time within which the expected status must occur so that the expectation can be considered satisfied. An Expectation is:
- *generated if* $\rho(\mathcal{W}^t) = true$
- *satisfied if* $e_s(\mathcal{W}^{t+\Delta t}) = true$

[1] In this paper, we give a simple role to the expectations, but we conceived them to be employed in most complex reasoning

3.3 Reasoning on Social Norms

The following algorithms provide the reasoner for a social robot for deciding to comply with social norms according to its expectations. Algorithm 1 is the core of the reasoner. The triple of elements it works is: *the state of the world* $\mathcal{W}^t$, the set of social values the robots wants to pursue represented by a set of qualitative goals ($\mathcal{QG}$), and a set of social norms $\mathcal{N}$. The state of the world $\mathcal{W}^t$ may change during system execution because the robot may perform some actions, it may perceive environmental changes, or it may capture events deriving from human interactions. For each active qualitative goal, the set of the related active social norms are considered. The most simple case is the presence of a single norm (Step Ⓐ). In this case, a desired final state is created according to the desirable operators. Analogously, Step Ⓑ allows for creating a desired final state by merging the different state of affairs the norms discipline. Thus, an achievement goal is generated by starting from the quality goal for reaching the new desired final state. After pursuing such goal, the new state of the world is updated.

Algorithm 1: Follow Social Norms

```
Data: W^t, QG, N
foreach QualitGoal_i ∈ QG do
    QualitGoal_i ← ⟨t_ci, q_si, s_ci, f_ci⟩;
    if QualitGoal_i is active then
        N_i ← {n ∈ N : n = ⟨p, q_si, ρ, φ, d⟩ ∧ φ(W^t) = true};
        Ⓐif card{N_i} = 1 then
            n ← ⟨p, q_si, ρ, φ, d⟩ ;
            if d = Des then
                f_s ← ρ;
            if d = Undes then
                f_s ← ¬ρ;
            if d = Indiff then
                f_s ← ρ ∨ ¬ρ;
        Ⓑ if card{N_i} > 1 then
            f_s ← ∅ ;
            foreach n_h ∈ N_i do
                n_h ← ⟨p, qs_i, ρ_h, φ_h, d_h⟩;
                if d_h = Des then
                    f_s ← f_s ∧ ρ;
                if d_h = Undes then
                    f_s ← f_s ∧ ¬ρ;
                if d_h = Des then
                    f_s ← f_s ∧ (ρ ∨ ¬ρ);
        AchievGoal ← ⟨t_ci, f_s, f_ci⟩;
        pursue(AchievGoal);
        update(W^t, f_s);
```

Algorithm 2 allows for evaluating the expectations that following a given norm may raise. In this first implementation, we consider that a robot has a satisfaction threshold that is decreased each time its expectation has not been satisfied after the time to fulfil. When its satisfaction reaches its lowest value, the related quality goal is dropped.

Algorithm 2: Evaluate Expectations

Data: $\mathcal{W}^t, \mathcal{EXP}$
foreach $exp_k \in \mathcal{EXP}$ **do**
  $\langle n, e_{sk} \rangle \leftarrow exp_k$;
  $n \leftarrow \langle p, qs, \rho, \phi, d \rangle$;
  $QG \leftarrow \langle t_c, q_s, s_c, f_c \rangle$;
  if exp_k *is generated* **then**
    $initTimeQs \leftarrow getCurrentTime()$;
    $update(\mathcal{W}^t, SensorData)$;
  $satisfied \leftarrow evaluate(e_{sk}, \mathcal{W}^t)$;
  if $(currentTime - initTimeQs) > ttf \wedge satisfied = false$ **then**
    $ThresholdQs = ThresholdQs - 1$;
  if $ThresholdQs = 0$ **then**
    $update(\mathcal{W}^t, f_c)$;

We want to highlight that the robot, in this case, does not follow all norms that are underlying the dropped goal. For example, if a robot wants to be polite, it tries to follow the social norms related to politeness such as say hello, thank you, sorry, etc. When the robot says hello, it expects that people greet it. Analogously, if the robot helps someone, it expects that the other person say thank you. If its expectations are continuously unsatisfied, similarly it could not say sorry when bumping into someone. It loses the motivation to follow the correlated norms because it thinks in this case that politeness is not a social value for the community of people it is interacting.

4 Case Study: Modelling robot sociality

In this section, we provide a simple case study for describing how a social robot may behave in some situations that involve social norms. To model robot and human objectives, we use the goal model diagram [20] where goals can be analysed, from the perspective of an actor, by Boolean decomposition, Contribution analysis, and Means-end analysis. Decomposition is a ternary relationship which defines a generic boolean decomposition of a root goal into sub-goals, that can be an AND or an OR decomposition. Contribution analysis identifies goals that can contribute positively or negatively towards the fulfilment of other goals. Finally, the Means-end relationship is also a ternary relationship defined among an Actor a goal and a task, representing the means which will be able to satisfy that goal. Practically, it provides the operationalisation of the goal.

The goal diagram showed in Fig. 1 represents the goal model related to a robot and human and their relations for the proposed case study. In the scenario under study, a robot has to go to a postal office for sending something on behalf of its owner. Thus, the robot has to reach the office, wait its turn sitting if there is a free chair, then talk with the postal employee for sending its item. In such scenario, some social norms regarding public behaviour have been made known to the robot. Such as i) *It is desirable to kindly greets when you meet someone*; ii) *It is desirable to say "I'm sorry" if you hit or bump into someone by accident*; iii) *It is desirable to be kind to the elderly, giving up your seat.* Thus, besides its functional objectives related to the physical tasks the robot is involved, a

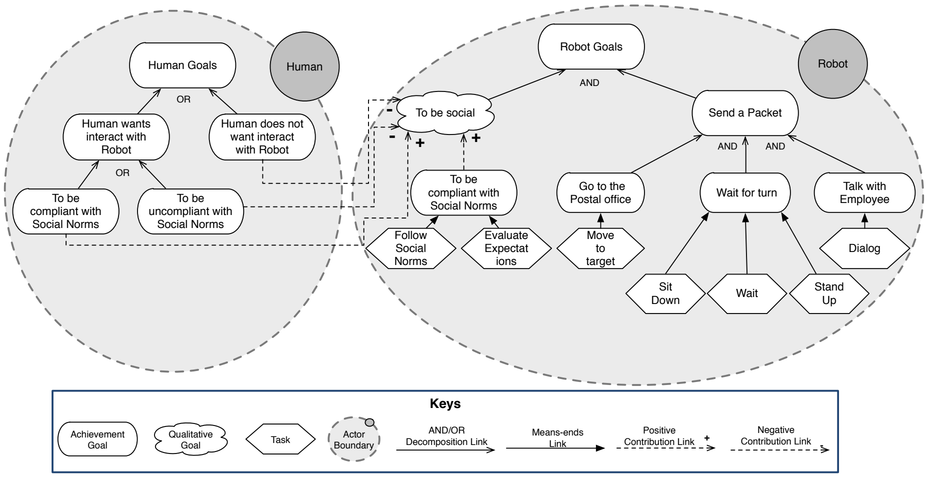


Fig. 1: Goal Model

qualitative goal represents the interest of the robot to be social. A goal that can contribute positively to reach this qualitative goal is to be compliant with social norms. Such goal is reached by accomplishing two tasks: follow social norms [2] and validate its expectations. Conversely, from the perspective of an individual, (s)he may have the interest to interact or not with the robot. In the latter case, the lacking of interactions negatively contributes to the fulfilment of the robot's qualitative goal because it lacks the fundamental requirement for being social that are the interactions. Conversely, an individual can have the interest to interact with a robot, but he/she can choose to be compliant with social norms or not thus violating the expectation of the robot. An individual that behaves conforming to the social norms favourites the robot's sociality because they satisfy the generated expectations of the robot. From the robot's perspective, the satisfaction of its expectations is a measure of the appropriateness of its behaviour. Conversely, individuals that do not follow social norms, thus not satisfying the expectations of the robot, could weaken the robot's belief about the appropriateness of the adopted social behaviour and cause him to change his attitude. In the initial implementation presented in this work, the robot suspends its interest in pursuing the social value underlying the disregarded social norms. According to our formalisation, the goal "*To be social*" and the above norms can be represented as follows [3].

G_1: `qualitative_goal(condition(want(be_social)),state(be_social),`
`condition(), condition(`$\neg$` want(be_social))`

[2] This is a generic task meaning that according to the social norms the appropriate task will be performed. In our case study, for example, the robot has to be able to greet, apologize and standing up for giving its seat.

[3] For space concern, we avoid to represent the achievement goals that are not relevant for understanding the case study.

N_1: `norm(position(_),state(be_social),state(greet),condition(is(person)), type(desirable))`

N_2: `norm(position(_),state(be_social), state(sorry),condition(bumped( person),type(desirable))`

N_3: `norm(position(_),state(be_social), state(standup),condition( and([is(person,old),state(sitted)])),type(desirable))`

The proposed case study has been tested in Choreographe by using a simulated Nao Robot. As we can see in Fig.2, the behaviour of the robot is not described using a predefined work-flow as it is commonly done in the Choregraphe environment. All the possible concrete tasks a robot may perform are linked to a normative reasoning component. Such component defines the behaviour of the robot according to its goals, thus deciding what tasks to be performed according to the specific circumstances the robot is working. As we can see, we implemented not only a set of concrete tasks the robot may use for satisfying its achievement goals, but also a set of tasks the robot may perform to be compliant with the previous set of social norms. Moreover, we developed a simple graphical interface to simulate some events such as met or bumped into someone or some conditions of the environment such as "there is a free chair." Such elements are perceived as beliefs by the robot which updates its knowledge about the state of the world. Each simulated scenario starts under the same conditions: the robot wants to send a packet and it wants to be social, W^0={`want(be_social),want(send,packet)`}. The expectations of the robot are met by the gratitude and politeness shown by the human. Each scenario presents three sections: a brief description of the scenario, the initial behaviour, the robot plan for reaching its achievement goals, and a description of the dynamic execution of the scenario.

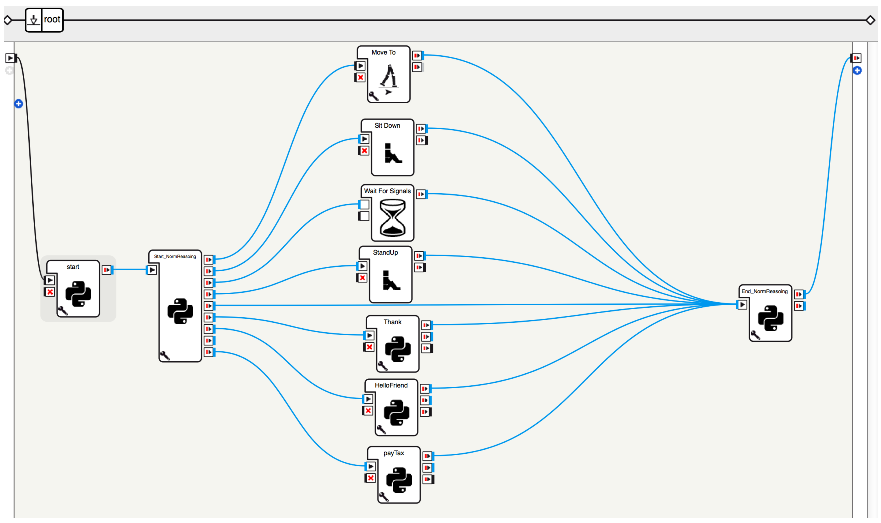
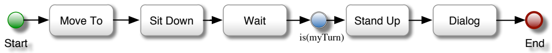


Fig. 2: Choreographe Components

SCENARIO 1

Description - In this scenario the robot arrives at the office, sees a free chair then it sits down and waits for its turn, after that it talks to the clerk. Such scenario shows the most simple situation where there are no applicable norms. Thus, any change occurs in the normal behaviour of the robot. The robot pursues its triggered achievement goals by following its initial planned behaviour.

Initial Behaviour



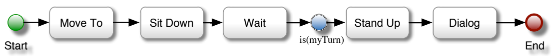
Execution

```
Task: Move to postal office
Task: Sit Down
Task:Wait
Expected Event: is(MyTurn)
Task: Stand up
Task: Dialog
```

SCENARIO 2

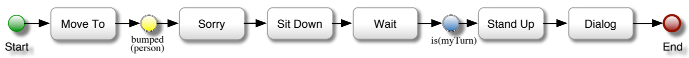
Description - In such scenario, the robot, moving to the postal office, bumps into a person. Such event triggers the norms N_2. Thus the robot changes its planned behaviour by adding the task of apologising. Then, the behaviour continues as in the previous scenario.

Initial Behaviour



Execution

```
Task: Move to postal office
Unexpected Event: bumped(person)
Applicable norm: norm(position(_),state(be_social), state(sorry),condition(
                 bumped(person),type(desirable))
Update Behaviour:
```

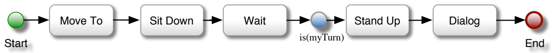



```
Task: Say Sorry
Unexpected Event: said(person,"Not at all")
Task: Wait
Expected Event: is(MyTurn)
Task: Dialog
```

SCENARIO 3

Description - In such scenario, the robot arrives at the postal office, sees a free chair then it sits down. An older adult arrives at the postal office. The robot changes its plan by following the norm N_3. Thus, it stands up and waits for its turn.

Initial Behaviour



Execution

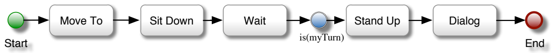
```
Task: Move to postal office
Task: Sit Down
Unexpected Event: is(person,old)
Applicable norm: norm(position(_),state(be_social), state(standup),condition(
                  and([is(person,old),state(sitted)])),type(desirable))
Update Behaviour:
```


```
Task: Stand up
Unexpected Event: said(person,"Thanks")
Task: Wait
Expected Event: is(MyTurn)
Task: Dialog
```

SCENARIO 4

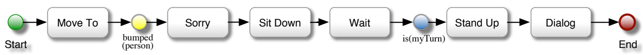
Description - In such scenario, the robot, moving to the postal office, bumps into a person. Such event triggers the norms N_2. Thus the robot changes its planned behaviour by adding the task of apologising. Then, it sees a free chair, and it sits down. An older adult arrives at the postal office. The robot changes its plan again by following the norm N_3. Thus it stands up and waits for its turn.

Initial Behaviour



Execution

```
Task: Move to postal office
Unexpected Event: bumped(person)
Applicable norm: norm(position(_),state(be_social), state(sorry),condition(
                  bumped(person),type(desirable))
Update Behaviour:
```




```
Task: Say Sorry
Unexpected Event: said(person,"Not at all")
Task: Sit Down
Unexpected Event: is(person,old)
Applicable norm: norm(position(_),state(be_social), state(standup),condition(
                  and([is(person,old),state(sitting)])),type(desirable))
Update Behaviour:
```


```
Task: Stand up
```

```
Unexpected Event: said(person,"Thanks")
Task: Wait
Expected Event: is(MyTurn)
Task: Dialog
```

In this section, we presented a simple case study for showing operatively the proposed approach. In particular, we want to highlight the flexibility of the approach. As we see, it is not necessary to define all the possible plans the robot may perform for managing all the possible situations. Indeed we do not implement if-then rules, but we provide the robot with the ability to reason about a mental concept such as norm and expectations. Thus, it can manage unexpected events that the robot does not consider in its initial plan.

5 Conclusions

Social robots interact with humans for performing specific tasks. Implementing social capabilities, such as behave following the social norms prescribed by the community, improves the social desirability of the robot. In this work, we propose a normative approach that allows exploiting the advantages of goal modelling to make social robots able to reason about dynamic situations pro-actively. In particular, we defined social norms by introducing desirability operators for representing preferences about acceptable behaviours and the expectations as a new mental concept a robot sees as a motivator for pursuing social values that we model using quality goals. Moreover, we have illustrated some scenarios about how the robot behaves in some situations that involve social norms, showing the flexibility of the approach to managing unexpected events.

As next step, we are working on allowing the robot to perform more complex evaluation about the expectations and how the robot may change its behaviour adaptively according to its expectations.

References

1. Jean-Léon Beauvois. Judgment norms, social utility, and individualism. *A sociocognitive approach to social norms*, pages 123–147, 2003.
2. Nicole Dubois and Jean-Léon Beauvois. Normativeness and individualism. *European Journal of Social Psychology*, 35(1):123–146, 2005.
3. N Dubois. Theory of the social value of persons applied to organizations: Typologies of good leaders and recruitment. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 60(4):255–266, 2010.
4. Nicolas Sommet, Alain Quiamzade, and Fabrizio Butera. How would pyrrho have been socially valued? social desirability and social utility of conflict regulation. *International Review of Social Psychology*, 30(1), 2017.
5. Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166, 2003.
6. John R Searle. *The construction of social reality.* Simon and Schuster, 1995.
7. David Lewis. *Convention: A philosophical study.* John Wiley & Sons, 2008.

8. Cristina Bicchieri and Ryan Muldoon. Social norms. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2014 edition, 2014.
9. Bertram F Malle, Matthias Scheutz, and Joseph L Austerweil. Networks of social and moral norms in human and robot agents. In *A world with robots*, pages 3–17. Springer, 2017.
10. Patrizia Ribino, Carmelo Lodato, and Massimo Cossentino. Modeling business rules compliance for goal-oriented business processes. In *Workshop on Enterprise and Organizational Modeling and Simulation*, pages 83–99. Springer, 2017.
11. Aldebaran Robotics. Nao robot, 2015.
12. Fabio Maria Carlucci, Lorenzo Nardi, Luca Iocchi, and Daniele Nardi. Explicit representation of social norms for social robots. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 4191–4196. IEEE, 2015.
13. What Social Robots Can, Should Do J Seibt, et al. Making place for social norms in the design of human-robot interaction. *What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016*, 290:303, 2016.
14. Ladislau Bölöni, Taranjeet Singh Bhatia, Saad Ahmad Khan, Jonathan Streater, and Stephen M Fiore. Towards a computational model of social norms. *arXiv preprint arXiv:1801.05796*, 2018.
15. Mehdi Dastani, M Birna Van Riemsdijk, and Michael Winikoff. Rich goal types in agent programming. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 405–412. International Foundation for Autonomous Agents and Multiagent Systems, 2011.
16. Geoffrey Brennan, Lina Eriksson, Robert E Goodin, and Nicholas Southwood. *Explaining norms*. Oxford University Press, 2013.
17. Carole Adam, Benoit Gaudou, Andreas Herzig, and Dominique Longin. Occs emotions: a formalization in a bdi logic. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, pages 24–32. Springer, 2006.
18. Anand S Rao. Agentspeak (l): Bdi agents speak out in a logical computable language. In *European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, pages 42–55. Springer, 1996.
19. Rafael H Bordini, Jomi Fred Hübner, and Michael Wooldridge. *Programming multi-agent systems in AgentSpeak using Jason*, volume 8. John Wiley & Sons, 2007.
20. Mirko Morandini, Fabiano Dalpiaz, Cu Duy Nguyen, and Alberto Siena. The tropos software engineering methodology. In *Handbook on Agent-Oriented Design Processes*, pages 463–490. Springer, 2014.