Pérula-Martínez, Raúl; Castro-González, Álvaro; Malfaz, María; Alonso-Martín, Fernando; Salichs, Miguel A. (2019). Bioinspired decision-making for a socially interactive robot, *Cognitive Systems Research*, 54, pp.: 287-301.

DOI: https://doi.org/10.1016/j.cogsys.2018.10.028

Bioinspired Decision-Making for a Socially Interactive Robot

Raúl Pérula-Martínez[a], Álvaro Castro-González[a,*], Maria Malfaz[a], Fernando Alonso-Martín[a], Miguel A. Salichs[a]

[a]*All authors are with the Department of Systems Engineering and Automation, Universidad Carlos III de Madrid, Spain.*

Abstract

Nowadays, robots and humans coexist in real settings where robots need to interact autonomously making their own decisions. Many applications require that robots adapt their behavior to different users and remember each user's preferences to engage them in the interaction. To this end, we propose a decision making system for social robots that drives their actions taking into account the user and the robot's state. This system is based on bio-inspired concepts, such as motivations, drives and wellbeing, that facilitate the rise of natural behaviors to ease the acceptance of the robot by the users. The system has been designed to promote the human-robot interaction by using drives and motivations related with social aspects, such as the users' satisfaction or the need of social interaction. Furthermore, the changes of state produced by the users' exogenous actions have been modeled as transitional states that are considered when the next robot's action has to be selected. Our system has been evaluated considering two different user profiles. In the proposed system, user's preferences are considered and alter the homeostatic process that controls the decision making system. As a result, using reinforcement learning algorithms and considering the robot's wellbeing as the reward function, the social robot Mini has learned from scratch two different policies of action, one for each user,

*Corresponding author

Email addresses: `rperula@ing.uc3m.es` (Raúl Pérula-Martínez), `acgonzal@ing.uc3m.es` (Álvaro Castro-González), `mmalfaz@ing.uc3m.es` (Maria Malfaz), `famartin@ing.uc3m.es` (Fernando Alonso-Martín), `salichs@ing.uc3m.es` (Miguel A. Salichs)

that fit the users' preferences. The robot learned behaviors that maximize its wellbeing as well as keep the users engaged in the interactions.

Keywords: Decision making system, autonomous robots, human-robot interaction, learning behaviors, artificial motivations.

1. Introduction

Recently, robots have been moved out of controlled environments (such as laboratories or production lines) to be introduced in more friendly conditions. In the last few years, a number of social robotic platforms have been developed with the capability of exhibiting social behaviors and collaborating with non-expert users in diverse environments (e.g. homes [1, 2], schools [3, 4], offices [5, 6, 7], hospitals [8, 9], or museums [10, 11]).

Social robots aim at interacting socially and communicate with humans following the behavioral norms expected by the people they interact with [12]. That is, since they are designed to live among humans, social robots should, for example, greet when they meet someone, or maintain a certain distance from their interlocutor while interacting. Real scenarios, particularly those involving human-robot interaction (HRI), are unpredictable and change continually. For instance, when a person is talking to a robot, unexpected events can happen; for example, the person can leave the conversation at any moment, (s)he changes the topic, or a new interlocutor arrives. This requires social robots to adapt their behavior to the environment and to make their own decisions.

Many researchers have focused their works on the adaptation of the robots' behaviors to unexpected events happening in the surroundings of the robot. For example, mobile robots are able to avoid unexpected obstacles encountered in their paths [13], or robots that grasp objects can deal with different forms and poses [14]. In the case of social robots, humans are now part of their environment, so they must be able to adapt their behavior to the people's unpredictable reactions. In this paper, we focus on the adaptation of the behavior of the social robot Mini to different kind of users while interacting. The goal is to provide the social robot with the capabilities to make decisions autonomously taking into account the particularities of each user, such as different reactions or preferences. In this work we consider the users' preferences in relation to the actions executed by the robot.

Moreover, in social robotics it is crucial to have robots that exhibit natural behaviors in order to ease its acceptability. In this context, we consider natural behaviors as those that can be observed in living beings such as animals or even humans. One way to achieve these behaviors is taking inspiration from nature, so we have employed a bio-inspired decision making system (from now on DMS) that includes motivations, drives, and wellbeing. This DMS drives the robot's actions in order to obtain well-accepted behaviors depending on the user. Particularly, in this work, we aim at achieving a social robot which interacts autonomously with different users, one at a time, and the robot's behavior is adapted to each user's preferences. Users' preferences have been incorporated in a homeostatic system as a robot's motivation and a drive, or need. This motivation and drive are not related to the robot itself (as most researcher do) but to an external agent (a user). This is a new approach for seeking the satisfaction of the user when interacting with the robot. This can be seen as a form of cognitive empathy where the robot reacts to the preferences of each user [15, 16]. This is important because, according to several researchers [17, 18], in order to achieve social interactions, empathy is one of the prerequisites. Moreover, for this adaptation, we have considered unexpected human actions that may occur at any time and we have modeled their effects as time-based states.

The rest of the paper is structured as follows. Section 2 reviews the most relevant contributions on bioinspired DMS that have been applied to social robots. After, Section 3 presents the DMS proposed in this article where motivations and drives lead the robot's behavior. The scenario where this DMS has been evaluated is described in Section 4. The results of the evaluation are commented on in Section 5 and, finally, the paper is concluded in Section 6.

2. Related works

Decision making in robotics is closely tied to answer two questions: what action does the robot have to execute? And when does it have to execute it? That is, a DMS selects the most appropriate action the robot has to perform at each moment.

Action selection has been extensively studied in robotics from decades. Since the late 80s, researchers looked for systems that combined goal-directed tasks and reactivity to anticipate changes in the environment [19]. In 1991, Brooks suggested robots that were able to select an action based exclusively on the changes of their surroundings [20], i.e. the behaviors exhibited by the robots where completely reactive. On the other hand, a few years later, researchers in behavioral psychology and artificial intelligence proposed that a behavior system needs to be composed by three types of elements: reactive, deliberative and reflective [21, 22, 23].

Other researchers took inspiration from the living beings and considered the *homeostatic drive* theory [24]. According to Cannon, *homeostasis* means maintaining a stable internal state [25]. According to the homeostatic drive theory, drive is an error signal that represents a deficit and the agent/animal acts to reduce the deficit and maintain an internal equilibrium. Drives evolve from a low value (or a satiated drive) to a high value when the deficit is very severe. Each drive is related to a motivation that leads the actions of the agent. Motivations compete to become the dominant motivation. Depending on the dominant motivation, the animal (or robot) selects the action or behavior to execute; for example, when an animal is hungry, the animal is motivated to eat so the animal consumes food and the drive hunger is reduced or satiated.

Some animal behaviors can be explained by the homeostasis theory so researchers have taken inspiration from it to obtain natural robot behaviors. When applying this theory to artificial agents, each agent has certain internal needs, such as hunger, companion or fun, which have to be kept within certain ranges to achieve the internal stability. When one or more of these internal needs are not satiated, a motivation urges the agent to act in order to satiate it.

There are several *homeostatic-based* architectures in the field of robotics that deserve special attention. The first one was developed by Velasquez in the late 90s [26]. He proposed the *Cathexis* architecture where a network of behaviors, such as "smile" or "express something", compete for the control of the agent. Each behavior contains two components: the *expressive* component and the *experiential* component. The *expressive* component includes aspects like prototypical facial expression, body posture, and vocal expression. The *experiential* component considers the motivations that affect the drives, as well as the action tendency and readiness that are modeled by the behaviors. The selection of actions in this architecture is made by a competition among behaviors in order to obtain the control of the agent: the behavior with highest value becomes the active behavior. This value is calculated from the motivations and a wide variety of external stimuli.

In 2000, Arkin et al. presented a bio-inspired model of the praying mantis that was applied to a robotic system [27]. In this model, there are three internal variables called motivational variables: fear (associated with predator avoidance), hunger (related to prey acquisition), and sex-drive (mating related). Each one of these variables is associated with a behavior that is enabled when the associated variable is the highest one. The enabled behavior is executed if a certain external stimulus is present. Otherwise, the next behavior with the highest motivation is evaluated.

Two years later, Arkin et al. studied the role of ethological and emotional models as the basis for an architecture that includes a behavior system for Sony's robot AIBO [28]. The mechanism of action selection in Arkin's architecture is based on evaluating both external and ongoing internal drives. They employed the "homeostasis regulation rule" where internal variables are specified and maintained within proper ranges. Behavioral actions and changes in the environment produced changes in the internal variables. In this architecture, the regulation of the internal variables was used as a motivational drive signal for selecting the behavior to be executed by the robot.

Later, Stoytchev and Arkin extended this work by considering circadian rithms in the evolution of the motivational variables [29]. In that work, four motivations (authors named them as curiosity, frustration, homesickness, and anger) changed their values based on different time-varying functions.

In 2003, Cañamero considered motivational states, e.g.*hunger* or *social attachment*, as internal drives or needs which were related to the survival of the agent [30]. Then, Cañamero proposed to use motivations, such as *curiosity* or *fatigue*, driven by the internal needs, that urged the agent to act. Motivations competed among themselves and the one with the highest value executed a behavior that contributed to satisfy the most urgent need(s) [31, 32].

In 2004, Breazeal designed a behavior system for the social robot Kismet. The interaction between the robot and the user is guided and inspired by that which occurs between a human infant with its caregiver. Kismet takes the infant role and the user is its caregiver [33]. Breazeal proposed that, in general, an animal can only pursue one behavior at a time. Therefore, each behavior is viewed as a self-interested goal-directed entity that competes against other behaviors for controlling the agent. Moreover, each behavior determines its own degree of relevance by taking into account the agent's internal motivational state and its perceived environment.

In the same year, Parisi focused on the importance of considering the internal elements of organisms when creating robots that are aimed at exhibiting natural behaviors[34]. In his simulations, Parisi considered physiological needs, such as food and water.

More recently, in 2013, Vouloutsi et al. proposed the Experimental Functional Android Assistant (EFAA) that contained multiple drives to display social competence and behaviors that promoted the HRI [35]. The EFAA was endowed with a repertoire of actions which were executed depending on the android's goal. These goals depended on the drives and each drive aimed at satisfying one goal.

Recently, Cao et al. [36] presented a *homeostatic* system adapted to the

HRI field. They proposed a hybrid concept for the behavior decision-making process which combines hierarchical (actions are linked to drives that compete to select the next action) and parallel (a set of actions is paired with each drive which has a priority and some preconditions that determine when to execute it) approaches. The robot behavior was selected based on external stimuli and drives, and the homeostatic system was used for triggering different artificial emotions too. The system was applied to an HRI scenario where a therapist and a Nao robot interacted with children. In a similar way, we propose to apply a homeostatic system to a real HRI scenario to lead the robot's behavior but taking into account the singularities of the individuals that interact with the robot.

Robot adaptation can happen at different levels. For example, Gomez-Donoso et al. [37] have developed a robotic system that adapts how patients with different capabilities interact with the robot during the therapy. The adaptation is conducted in a previous phase and it is conducted with the assistance of a therapist. The exercises realized during the therapy are predefined by the experts. Then, in this case, the adaptation is supervised by an external person and it is related to how the users communicate with the robot and the exercises conducted.

Homeostatic-based DMSs have been used to adapt the robot's behavior to social aspects. In particular, Hieida et al. briefly combines a homeostasis process with reinforcement learning [38]. The robot's affective state is part of the homeostasis process which forms the reward signal in the learning process.

Several works have shown the importance of the adaptation of the robots' behaviors when they have to interact with people [39, 40]. Recently, Rossi et al. [41] have presented a survey where the robot behavioral adaptation is classified in physical, cognitive, and social. Focusing on the social adaptation, social signals coming from the people interacting with the robot should be considered by the robot. Vinciarelli and Pentland [42] considered social signals as observable behaviors that produce behavioral changes during the interaction. Based on this definition, one of these social signals is the user's preferences.

Users' preferences is a signal that has already been used in robots that are involved in HRI [43, 44]. Preferences are related with the emotional state. Thus, Tanevska et al. proposed a DMS where the iCub robot adapted its behavior to the person's preferences considering the user's emotional state [45]. Tanevska et al. used one Markov chain per each robot's action modeling the probabilities of transition among three users' emotional states (neutral, bored, and interested). These states were determined by an image processing algorithm that used images of the user's face. However, authors did not evaluate the adaptation of the robot but the perception module.

Other researchers have presented works where the robot adapts its behavior to the people the robot interacts with. This is the case of the study made by Ramachandran et al. [46], where a social humanoid robot is used to tutor children in one-on-one interactions. They outlined an architecture in which the robot used reinforcement learning to adapt the difficulty of tutoring exercises of arithmetic problems to each child. The engagement level and the learning gains were used for the reward signal. Similarly, thanks to our DMS, the robot customizes its behavior to different users.

In addition to the users' preferences, it is important to consider unexpected events that could happen during the HRI. Recently, Görür et al. considered unexpected human behaviors in a collaborative task where a human operator and a robot work together in a factory [47]. In the work of Görür et al., authors used POMDPs to create an stochastic decision-making mechanism where the partially observed states are related to the operator's intentions and the robot decides when to assist her. In this model the actual action of the operator and its consequences are not considered in the DMS, but the system considers a model of the humans that helps to predict her future actions and the robot acts accordingly. In that work, the evaluation is conducted in a virtual environment where the robot and the operator are simulated. In this line, we are interested in the effects of the unexpected human actions, rather than just user's intentions, on the robot in real settings.

Based on previous works [48, 49, 50], in this paper we propose to use a homeostatic approach to allow a robot to make decisions considering the person located around the robot. The main contribution of this work is twofold. First, in contrast to the previous work, the DMS is tailored to foster the HRI and adapt the robot's behavior to the user's preferences towards the robot's actions. This is achieved by including the user's preferences in our homeostatic-based DMS. Secondly, we consider unexpected human actions happening at any time during the interaction (for example, a user approaches the robot or a user interrupts the interaction). The effects of the user's actions are modeled as transitional states and considered in the learning mechanism that allows the autonomous adaptation of the robot to the user's behavior.

3. The Decision Making System

3.1. Drives and motivations in the homeostatic process

Following a similar approach to that presented in several of the works mentioned in section 2, our DMS includes a homeostatic process. The term homeostasis refers to a state of psychological equilibrium obtained when a drive has been reduced or eliminated. Thus, the robot has certain needs or **drives** that should remain at their lowest values. When a drive deviates from its minimal value, a **motivation** arises to urge the robot to take action and overcome the deficiency. For example, when a predator is hungry, its drive related to hunger is very high and the motivation to eat makes it to prey on other animals.

Sometimes, motivations arise because of perceptual stimuli rather than internal causes. For instance, when a kid sees chocolate, whether she/he is very hungry or not, she is motivated to eat it. These perceptions from the environment are called **external stimuli** and influence the robot's decision making process by altering the motivations to behave in one way or another. This is inspired by the Behavioral Theory [51] where Hull proposed the idea that motivation is determined by two factors: the first factor is the drive; and the second one is the incentive, that is, the presence of an external stimulus that predicts the future reduction of the need.

In our approach, drives are considered as the robot's needs and their ideal value is **zero** (satisfied). Each drive evolves automatically increasing its value until the saturation value is reached or reduced due to an action (e.g. the predator preys on other animals). When a drive is satiated due to a robot's action, the drive remains satiated (i.e. with a value of zero) for a while before it starts to evolve again. This is called the satisfaction time. According to the Behavioral Theory [51], the intensities of the robot's motivations are modeled as a function of its drives and some external stimuli. Depending on the level of a drive, the stimulus needed to trigger a motivation can be intense or weak. Therefore, the value of the motivations is calculated as shown in Equation 1.

$$\begin{aligned} &If \quad D_i < L_d \quad then \quad M_i = 0 \\ &If \quad D_i \geq L_d \quad then \quad M_i = D_i + w_i \end{aligned} \tag{1}$$

where M_i is a particular motivation, D_i is its related drive, w_i corresponds to the related external stimuli, and L_d is the activation level. The activation level is defined as a threshold that makes the motivation relevant just after the related drive has reached a certain value.

Motivations are competing continuously among themselves for being the dominant one. The motivation with the highest value is considered the **dominant motivation** and it leads the robot's actions. According to Equation 1, motivations whose drives are below their activation levels will not be able to lead the robot's behavior.

3.2. The robot's state

In our approach, the robot selects the next action to be executed depending on its current state. The robot's state has been defined as the combination of the internal and external states. The internal state is determined by the dominant motivation and the external state is related to the objects the robot can interact with. Continuing with the example of the predator, its actions could be different if the predator is motivated to eat or to rest (internal state), and also if the predator is alone or accompanied by other predators (external state). Mathematically, the state of the robot $s \in S$ is represented in Equation 2.

$$S = S_{internal} \times S_{external} \tag{2}$$

In relation to the external state, the robot might interact with multiple objects. In this work, we aim at achieving a social robot which interacts autonomously with different users one at a time. Therefore, the objects the robot is able to interact with are the different users who communicate with the robot and its external state is related to the state of the users ($S_{external} = S_{user}$).

As an example, consider the situation where $user_A$ is interacting with the robot and, after a long time of activity, the robot's need to *relax* is very high making the motivation to *rest* become the dominant one. In this situation, the state of the robot is presented in Equation 3.

$$\begin{aligned} S &= S_{internal} \times S_{external} = \\ &= S_{dominant\,motivation} \times S_{user_A} = \\ &= rest \times user_A(interacting) \end{aligned} \tag{3}$$

3.3. Learning a policy of actions

As already stated in section 1, social robots must be autonomous in order to exhibit a natural behavior. This implies that these robots have to make decisions based on their state and their repertoire of actions. The robot's action selection policy maps states and actions. This policy can be predefined by the roboticists or it can be learned by the robot. In this work, our goal is to have a robot that is able to adapt its behavior autonomously to different users and, at the same time, to maintain its needs within an acceptable range. Since the way each user behaves is unknown and very different from one to another, a predefined policy is unpractical. Then, we have developed a mechanism that allows the robot to learn from scratch the best action to execute depending on its most urgent need (the internal state) and its state in relation to the user present at each moment (the external state).

In line with our previous work [48], we use reinforcement learning (RL) as the unsupervised learning technique to find out the robot's best actions for each user in different situations. RL is inspired by the behaviorist psychology and concerned with how agents ought to take actions in an environment so as to maximize a reward. The goal of RL is then to maximize the total expected reward. RL differs from standard *supervised learning* in that correct input/output pairs are never presented, but the agent learns from direct interaction with the environment.

In particular, the learning algorithm included in our DMS is the Q-Learning algorithm [52]. This algorithm consists in an agent that is in a state s_t and executes an action a. When the action is done, the agent has transited to a new state s_{t+1} and receives a reward r. The learning algorithm updates the Q-value for that pair (s_t, a) according to the obtained reward r and the new state s_{t+1}. These *Q-values* represent how good it is to execute a particular action when the agent is at a particular state. Q-values are defined as the expected reward when executing an action a in the state s $(Q(s, a))$.

In our case, the robot learns the best action for each user individually so our DMS considers different Q-values for each user. The Q-values for a particular user are updated according to Equation 4.

$$\begin{aligned} Q_{user_i}(s,a) = (1-\alpha) * Q_{user_i}(s,a) + \alpha * (r + \gamma * V_{user_i}(s')) \\ s, s' \in S_{user_i}; a \in A_{user_i} \end{aligned} \tag{4}$$

S_{user_i} is the set of states in relation to $user_i$ and A_{user_i} is the set of actions related to $user_i$. $V_{user_i}(s')$ is the value of the state s' and it is the best reward the robot expects from state s'. $V_{user_i}(s')$ is calculated as shown in Equation 5.

$$V_{user_i}(s') = \underset{a \in A_{user_i}}{max} (Q_{user_i}(s', a)) \tag{5}$$

Parameters γ, α and r are respectively the discount factor, the learning rate, and the reward.

The discount factor γ defines how much expected future rewards affect decision now. A high value of this parameter gives more importance to future rewards. On the contrary, a low value gives much more importance to the current reward.

On the other hand, the learning rate α controls the weight provided to the reward from the action made. This parameter gives more or less importance to the learned Q-values than new experiences. A low value implies that the robot is more conservative and therefore gives more importance to past experiences. On the contrary, a high value makes the agent values the most recent experience.

In relation to the reward, r, we have used the variation of the robot's wellbeing as the reinforcement received after the execution of every action. This approach was inspired by Gadanho [53] and it has already been used by the authors in prior works [48, 49, 50]. The robot's wellbeing (Wb) is related to its needs and it is computed as presented in Equation 6.

$$Wb = Wb_{ideal} - \sum_i D_i \tag{6}$$

Wb_{ideal} represents the ideal value of the wellbeing when all drives are satiated ($\sum_i \times D_i = 0$). It corresponds to the maximum robot's wellbeing or, in relative terms, the 100% of its value.

On the other hand, the higher the values of the drives, the lower the value of the wellbeing. In view of this definition, the reward is calculated as the variation of the robot's wellbeing before and after executing the action a (Equation 7).

$$reward = \Delta Wb_a = Wb_{after\,a} - Wb_{before\,a} \tag{7}$$

It is important to mention that the robot's actions have effects over the drives. Some actions reduce or satiate a drive (e.g. playing with a user reduces the need of *interaction*) but others can increase their values (e.g. playing a game increases the need to *rest*).

Consequently, when the robot executes an action that causes a significant drop in the value of a drive, this is reflected in an increase in the robot's wellbeing and therefore in a positive reward. This can be understood as executing that action from that state is a good decision. On the contrary, a negative reward is caused by an action that leads to a rise of the robot's drives and, correspondingly, it produces a reduction of its wellbeing.

4. Evaluation

4.1. The robotic platform

The robotic platform used in this work is Mini (Figure 1), a desktop social robot created to interact with people, in particular with elders suffering cognitive impairment [54]. Mini is able to conduct meaningful interactions in order to perform cognitive stimulation exercises or to play educational games.

Considering that Mini is a desktop robot, it cannot move around and therefore the potential HRI is limited by this aspect. Having this in mind, the location of the user in relation to the robot is a key aspect to consider while the user interacts with the robot.

Regarding the hardware components, Mini's head includes: two screens where animated expressive eyes are displayed, two RGB-LED cheeks, and a VU-meter-like mouth that illuminates according to the volume of the audio signal generated by the robot. Mini can move its head by means of a 2 DOF neck (pan and tilt) and it is endowed with two 1 DOF arms. Its torso includes a colorful LED-based heart that *beats* and changes its color. Furthermore, several touch sensors are located in the body to detect when and where the robot is touched. In addition, a microphone and two speakers are located in the belly to carry out verbal communication.

In the base, a depth camera (Kinect) eases the detection and identification of different users around the robot. The main computer and a data acquisition board are placed inside the base.

An external tablet is used to show videos, images, or menus during games or exercises. Finally, an external button is used to provide the push-to-talk functionality: when it is pressed, the automatic speech recognition module starts to process the input audio signal to extract the meaning of the user's utterances.

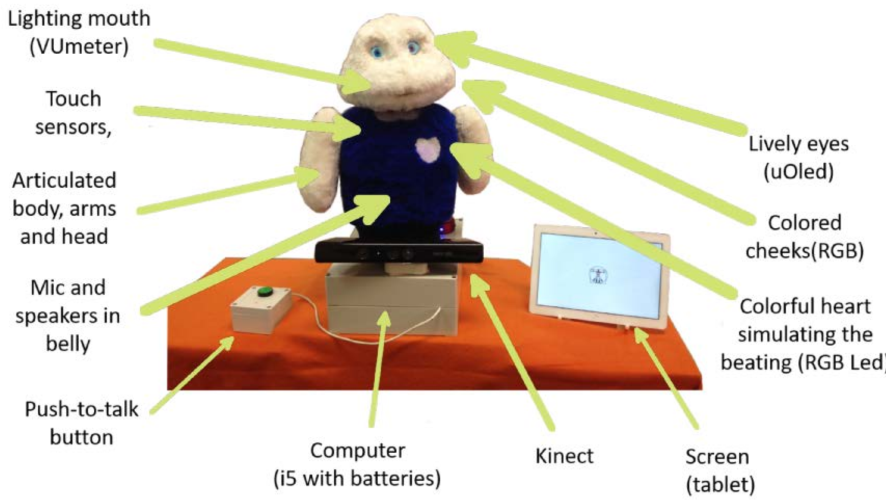


Figure 1: The social robot Mini and its hardware architecture.

4.2. Customization of the Decision Making System

In this study, we desire to have an interactive robot which aims at creating social bounds with different bystanders and engaging them in playing interactive games. Having this in mind, the robot's DMS has been tailored to achieve these goals.

It is important to remark that the configuration presented in this section is a design decision that will affect the robot's behavior. That means that other values or parameters would result in a robot showing different behaviors. Studying how the parameters of the DMS affect the robot's behavior is out of the scope of this work.

As already mentioned in Section 3, in our DMS, the state of the robot is a crucial concept that is represented as a tuple formed by the internal and the external states ($S_i \ \times \ S_e$).

4.2.1. The internal state S_i

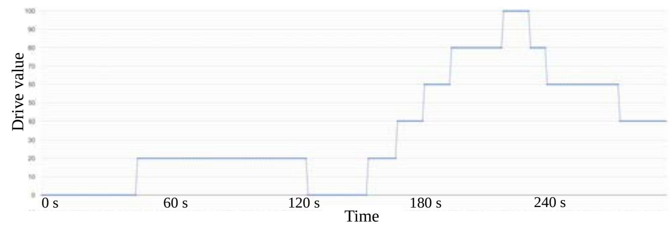
In this experiment we have defined three motivations: **please**, **relax**, and **socialize**; each one is associated with the following drives respectively: **user's satisfaction**, **rest**, and **interaction** (see Table 1). In contrast with some of the previous works presented in Section 2 [26, 28, 30, 31, 35], our drives are not related with physiological needs or survival (e.g. food, energy, or security). In line with Cao et al. [36], our drives are aimed at fostering the HRI.

Motivation	*Drive*
please	user's satisfaction
relax	rest
socialize	interaction

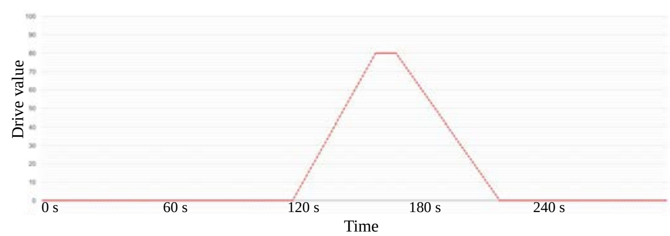
Table 1: Motivations and their associated drives.

In order to ease the engagement in the interaction, we consider that the user experience with the robot is very important. Thus, the motivation *please* has been designed to consider the user's enjoyment when interacting with Mini. The drive associated to this motivation, called *user's satisfaction*, decreases when the user is pleased and increases when she is disappointed; that is, the more she likes the robot, the lower the robot's need of satisfying the user (because she is already satisfied). Thus, the value of the *user's satisfaction* drive changes when the user shows its satisfaction or disappointment with the interaction (see the step-shaped plot in Figure 2a). Its saturation value (i.e. its maximum value) was set to 100 and, considering that it is the highest one among all drives, the associated motivation *please* is the most urgent one and Mini's primary goal is to keep its users satisfied.

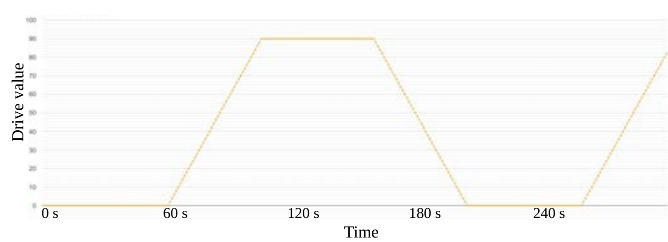
In the case of *relax*, this motivation has been included to avoid a hyperactive robot that never gets tired. This motivation helps to achieve a more natural behavior where a robot sometimes is proactive but, after a while, it needs to *rest*. In this case, the need to *rest* (its associated drive) increases when the robot is performing an action ($2.67\,points/10\,seconds$) and it decreases when it is idling



(a) User's satisfaction drive. When the step-shaped plot goes up, the user has enjoyed the interaction; when it goes down



(b) Rest drive. The drive increases when



(c) Interaction drive. When the robot is interacting with a user, this drive decreases; otherwise, it increases.

Figure 2: Evolution chart for the drives.

(see Figure 2b). This drive ranges from 0, its initial value, to 80, its saturation level, and the satisfaction time is 120 seconds. For this motivation, the potential HRI is considered as an external stimulus, an incentive, then, when a person is close enough to Mini, the value of the motivation *relax* rises 10 points.

We believe that the social bounds between the users and Mini will be established after several interactions. To foster these social interactions, the motivation *socialize* impels our robot to communicate and interact with people. The need of *interaction* grows when no one is interacting with the robot (at a rate of $4\,points/10seconds$) and, on the contrary, drops at the same rate while HRI is happening (see Figure 2c). For this drive, the saturation level has been established to 90 points, and the satisfaction time is set to 60 seconds. In addition, the presence of a user represents an external stimulus for the motivation of *socialize* and its value increases 10 points.

At the beginning of the scenario, all drives are satiated and consequently their initial value is 0.

The parameters used in the configuration of the internal state have been decided considering two criteria: first, our previous experience with this type of DMS in robots [48, 49, 50]; and second, the parameters have been adjusted to obtain a robot that experiences all possible situations for the internal state (i.e. dominant motivation). Before running the evaluation, we have empirically tested the DMS and observed the robot's behaviors using different values for the parameters. When Mini experienced all dominant motivations, those were the values selected for the evaluation.

4.2.2. The external state S_e

As previously described in Section 3, the external state is represented by the state of the robot in relation to all items. In this work, the robot is intended to interact with users. Therefore, S_e is composed by the state of Mini in relation to the people that Mini interacts with. We have limited the scenario to 1-by-1 interactions, which means that the robot interacts with one user at a time. Then, for each user (u_i) we have the following three basic robot's states:

- User is absent: u_i is not perceived by Mini. This is the default state for all users.
- User is near: the robot detects u_i in the surroundings.
- User is interacting: u_i is right in front of Mini facing it. In this state, we consider the user is interacting with the robot, or willing to.

Users are considered as autonomous agents that act by themselves and their actions can affect the robot's state. These actions, from the robot's perspective, are *exogenous actions*; that is, actions that are executed by other agents (or people in this approach) and cause changes in the robot's environment, but they cannot be controlled by the robot. For example, if a person turns off the light of the room where the robot is, this action alters the conditions where the robot is operating. These exogenous actions can trigger unexpected changes of the robot's state. For example, if the robot is alone in a room and a person enters, now the robot is accompanied but that change in the robot's state has not been due to a robot's action, but due to an action of that person. Considering exogenous actions by the robot's DMS is an open problem.

In this work, we have considered the effects of the exogenous actions. When the effect of an exogenous action is perceived, a transition to a time-based state is triggered; after a predefined time window, the system transits automatically to another state. In this approach we have considered the exogenous actions related to the user's displacements and we have ended up with the next four time-based robot's states in relation to each user:

- User is appearing: u_i has entered into the perception field of the robot.
- User is approaching: u_i was near Mini and has moved right in front of the robot.
- User is leaving: after interacting, u_i walks away from Mini.
- User is disappearing: u_i leaves the area where the robot is so the user exits the perception field of the robot.

Note that all transitional states are associated with user's actions. These states are active for a limited amount of time that was empirically set to 5 seconds.

Figure 3 shows all the external states considered in this experiment for each user.

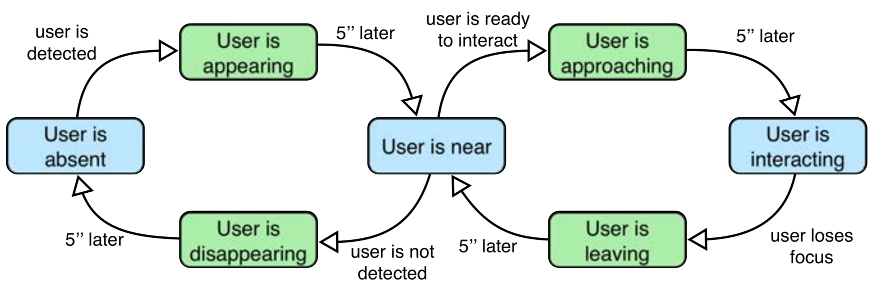


Figure 3: External states of the robot considered for each user.

4.2.3. Repertoire of actions

The scenario defined for this experiment is an educational game where different mathematical questions (according to different levels) are asked by the robot. Mini has been endowed with a predefined set of actions to engage users in this scenario. It is important to mention that Mini has the repertoire of actions but it does not know when to execute each one of them. The DMS proposed in this work first learns the best policy for each user and then it selects the proper action in every situation according to that policy.

The available robot **actions** for this scenario are:

- wait: Mini remains idle and says utterances like "I'm tired".
- ask an easy question: Mini asks an easy mathematical question.
- ask a medium question: this is equal to the previous one but the difficulty of the questions increases.
- ask a hard question: in this case, the robot asks the most difficult mathematical questions.

- attract attention: Mini tries to draw the users' attention by greeting them or using utterances like “Is anybody there?”, “Does anyone want to play with me?”, or “Are you leaving?”.

It must be said that after the user has answered a mathematical question, either verbally or through the tablet, Mini asks for the *user's satisfaction* displaying a message in the table ("Rate your satisfaction with the game”) and using a 3 star menu. A rating of 1 star means that the *user's satisfaction* is very low, a 2 star score represents a medium satisfaction, and the maximum satisfaction is represented by the 3 star rating.

As explained in Section 3, all actions cause some kind of effect. These effects alter the drives of the robot and consequently its wellbeing. The effects of all actions are summarized in Table 2. Notice that most of the actions raise 5 points the value of the *rest* drive. This effect represents the “effort” of executing an action and consequently the need of *rest* increases. In the case of the action *wait*, this drive decreases at the rate of 2.67 points every 10 seconds. The longer the robot waits, the longer it rests, and the lower need of *rest*.

In relation to the drive *user's satisfaction*, its value changes depending on how the user has enjoyed the interaction. As already explained, this is evaluated through a 3 star menu after the user's responds: if the user rates the interaction with 3 stars, the *user's satisfaction* drive is reduced by 10 points; if she gives 1 star, this drive increases 10 points; a 2 star rate does not change the drive. As we mentioned earlier, one of the robot's motivations is to *please* and this is achieved when the user enjoys the interaction with Mini.

Notice that the answer to the math questions (either easy, medium or hard) does not have an effect on the drives. That is, whether the user's answer is right or wrong is not relevant because it does not change the value of the drives.

Remember that the drive *interaction* changes depending on the state of the users (it rises when the user is interacting and it drops when she is not) so it is not affected by the robot's actions.

Action	**Drive: effect**
wait	rest: $-2.67\ points/10\ seconds$
ask easy question	user's satisfaction: -10 if 3-star rating user's satisfaction: +10 if 1-star rating rest: +5
ask medium question	user's satisfaction: -10 if 3-star rating user's satisfaction: +10 if 1-star rating rest: +5
ask difficult question	user's satisfaction: -10 if 3-star rating user's satisfaction: +10 if 1-star rating rest: +5
attract attention	rest: +5

Table 2: Actions-Effects relation.

4.3. The interactions

In order to evaluate the DMS, we considered two different user profiles (Section 4.4). Each user interacted with Mini during 5 hours, divided in 4 sessions: three 90 minute sessions and a final 30 minute session (see Figure 4).

In order to select the action to execute, we used the Boltzmann distribution [52]. This method uses a parameter called *temperature* (T) to balance between exploration and exploitation. A high values of T benefits the random selection of the actions, independently of the Q-value associated. On the contrary, a low value of T implies a reduction on the randomness so the action selected will be the one with the highest Q-value. Considering that the robot was learning from scratch, it needed time to learn the best actions in each situation. Then, the exploration phase was composed by the first 3 learning sessions (Session 1, Session 2 and Session 3 in Figure 4) where the robot selected the next action to execute randomly; thanks to a high value in the *temperature* parameter ($T = 100$), in these learning sessions, the Q-values barely influenced the action selection. In the exploitation phase, the final session (Session 4 in Figure 4), the *temperature* is drastically reduced ($T = 0.1$). Consequently, in this last session, the robot selected the actions according to the learned Q-values, that is, it exploited the acquired knowledge by following the learned policy of actions.

Following the approach used in our previous work [50], the learning rate (α in Equation 4) was reduced gradually from 0.3 in Session 1 to 0 in Session 4. This implies that during the last session, the Q-values are not updated any more and Mini exploited the learned values.

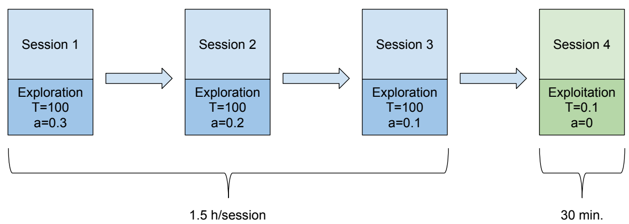


Figure 4: Sessions defined for the HRI experiments. 'a' corresponds to the learning rate and 'T' is the temperature factor that balances the exploitation and the exploration.

4.4. User profiles

Our DMS has been designed to adapt the robot's behavior to the user through to the interaction. To show it, we have considered 2 antagonistic user profiles to demonstrate how our system is able to adapt the robot's behavior to very different users. These profiles describe the behavior of two users when they are nearby the robot.

In this case, we considered the users' preferences towards the robots actions; in particular, the user profiles differ mainly on the preferences for the level of the mathematical questions. Then, when talking about the users' preferences, we refer to the different users' ratings of each game after they play with the robot.

Therefore, Mini will learn 2 policies of action, one for each user profile, using our DMS.

4.4.1. User profile 1

The user profile 1 (UP1) is a curious person that is attracted by the robot but it interacts with the robot sporadically. The user is not particularly attracted by a specific robot behavior, so the user does not prefer a behavior over the others.

When playing the quiz with Mini, she likes to answer correctly the maximum number of questions and she is not interested on challenging questions. This means that, when the user answers correctly a question, she rates higher her satisfaction. On the contrary, when she answers incorrectly, her ratings are lower.

4.4.2. User profile 2

The user profile 2 (UP2) represents a sociable person that is willing to interact with Mini as many times as possible. Thus, when Mini calls her attention, this user approaches the robot.

Furthermore, she likes challenging questions, despite she might not know the answer. Then, when facing challenging questions, her satisfaction increases. In contrast, she is very disappointed with easy questions and her satisfaction decreases.

5. Results

The evaluation of the DMS system has been conducted considering the learned policy of actions for each user profile and the robot's wellbeing during the exploitation phase. It is important to remember that, in the exploitation phase, the DMS selected the best action depending on the robot's state to maximize Mini's wellbeing. In the following, we present the results for each user profile.

5.1. User profile 1

Figure 5 represents the policy of actions learned for interacting with the UP1. That is, the best actions to be executed for each dominant motivation depending on the state of the robot in relation to the UP1.

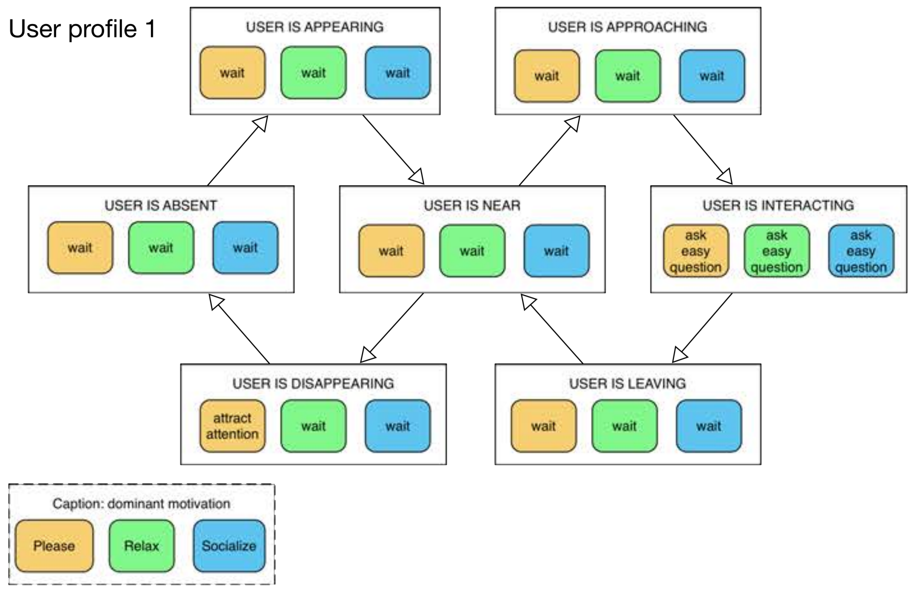


Figure 5: Learned policy for User Profile 1. White boxes represent the states related to the user and colored boxes show the best action to execute for the different dominant motivations (orange: please; green: relax; blue: socialize).

5.1.1. Dominant motivation: please

In relation to the *please* motivation, the robot has learned that the best action when interacting with UP1 is to ask easy questions because most of these questions are correctly answered and, when this has happened, UP1 has rated her experience higher. Let us recall that UP1 enjoyed to answer correctly the maximum number of questions and this is more likely when asking easy questions.

Considering that UP1 approaches the robot regardless of what the robot is doing, Mini has learned that, for the rest of external states, the best action is to wait. During the learning, Mini tried all possible actions in each external state but, since UP1 did not show any preference for any robot's behavior (as it is described where the UP1 is presented, Section 4.4.1), none of the actions was evaluated as very positive for the robot's wellbeing. The resulting best action, waiting, represented a small positive reward because it reduces the need of resting and consequently represents a positive variation of the wellbeing.

Focusing on the external state *user is disappearing*, the action with the highest Q-value is *attract attention*. This result may seem strange but, in comparison with the other states, this one was barely explored; this implies that presumably the robot did not have enough chances to learn the right action in this state. We believe that longer exploring sessions would have resulted in more chances to explore this state and the consequences of the actions from it, and likely in a different best action for that state.

5.1.2. Dominant motivation: relax

In the case of the *relax* motivation, predictably, the best action in most of the external states is to wait. In this situation, the effect of waiting is to reduce the drive *rest*, which is related to the dominant motivation *relax*, and it obtains a large positive variation of the Mini's wellbeing. Notice that just when the user is interacting, the learned action is to ask easy questions, instead of waiting. This can be explained if you consider two effects: (i) in this state, the drive *interaction* decreases; and (ii) the more right answers, the more UP1 likes the interaction and hence the *user's satisfaction* drive is reduced too. This double drop on the robot's needs represent a very large increment on its wellbeing, even higher than the reward obtained when the drive related to the dominant motivation is reduced.

5.1.3. Dominant motivation: socialize

The learned behavior when the highest motivation is *socialize* is the same as when the dominant motivation is *relax*. Again, this is a consequence of the unpredictable user's behavior: her reactions are unrelated to the robot's actions and the action the robot executes does not drag the user to the *interacting* state. The double reward obtained when asking easy questions while interacting is observed here too, even stronger. In this case, when *interaction* is the dominant motivation, the reward is higher when the user interacts longer with Mini because she is entertained answering successfully (this occurs more frequently with easy questions).

5.2. User profile 2

5.2.1. Dominant motivation: please

The learned policy for UP2 is presented in Figure 6. When *please* is the dominant motivation, the need of *user's satisfaction* is very high and the robot seeks the way to please the user. This is achieved when UP2 rates the interactions satisfactorily and, considering this profile, this happens when the robot asks her hard questions (see on Figure 6 that the best action while interacting when *please* is the dominant motivation is to ask hard questions).

Using the RL algorithm, the robot does not learn only the immediate action, but the sequence of actions to satiate a drive. This can be clearly observed with UP2 when *please* is the dominant motivation: the best action in most of the states is to attract the user's attention, which is how Mini can afterwards interact with the user. Notice that when UP2 is absent, this means that Mini can not perceive the presence of the user but it does not mean that the user cannot hear or see Mini. Actually, because of the way UP2 behaves, when UP2's state is absent or disappearing, Mini acts to attract her attention, UP2 perceives it, she is interested on Mini and approaches it. As a consequence to this behavior, the best action Mini can do in the subsequent states, appearing and near, is to wait since UP2 is going to interact in any case.

5.2.2. Dominant motivation: relax and socialize

As it happened with UP1, in the cases of the dominant motivation is *relax* or *socialize*, the learned behavior is the same for both inner states. It should be noted again that, when interacting, the robot's wellbeing increases due to a double reduction of the drives: first the *interaction* drive drops because Mini and the user are interacting, and second the UP2's favorite action is executed (ask hard question) resulting in the fall of the need of the *user's satisfaction*. In particular, in the case of the *relax* motivation, this two-sided reward is higher than that obtained by waiting, which would make more sense.

In the other states for UP2, waiting is the best action. This is obvious in the case of the *relax* motivation because it reduces the drive *rest* whose value should be very high since it is associated with the dominant motivation.

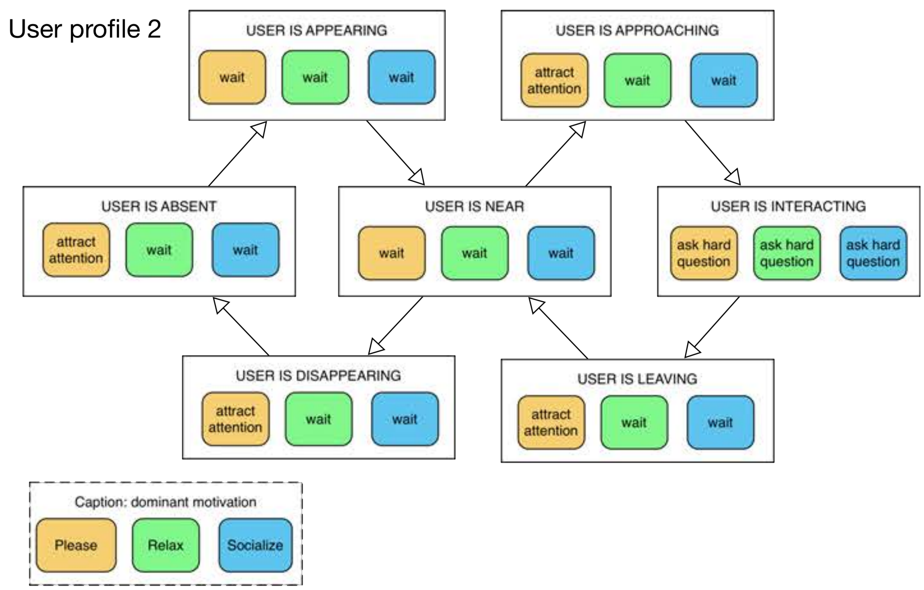


Figure 6: Learned policy for User Profile 2. White boxes represent the states related to the user and colored boxes show the best action to execute for the different dominant motivations (orange: please; green: relax; blue: socialize).

5.3. *Mini's Wellbeing*

During learning (exploration phase), Mini executed actions randomly to learn their consequences and, as a result, its wellbeing was adversely affected. On the contrary, during the exploitation phase, Mini used the learned policy to select the best actions (in terms of Mini's wellbeing) to be executed. To show how the DMS works after learning the policy of actions, in this section we analyze the robot's wellbeing during the 30 minute exploitation phase for UP1.

Figure 7 details the evolution of the dominant motivation, the UP1 state, the executed actions, and the robot's wellbeing. Initially, *please* is the dominant

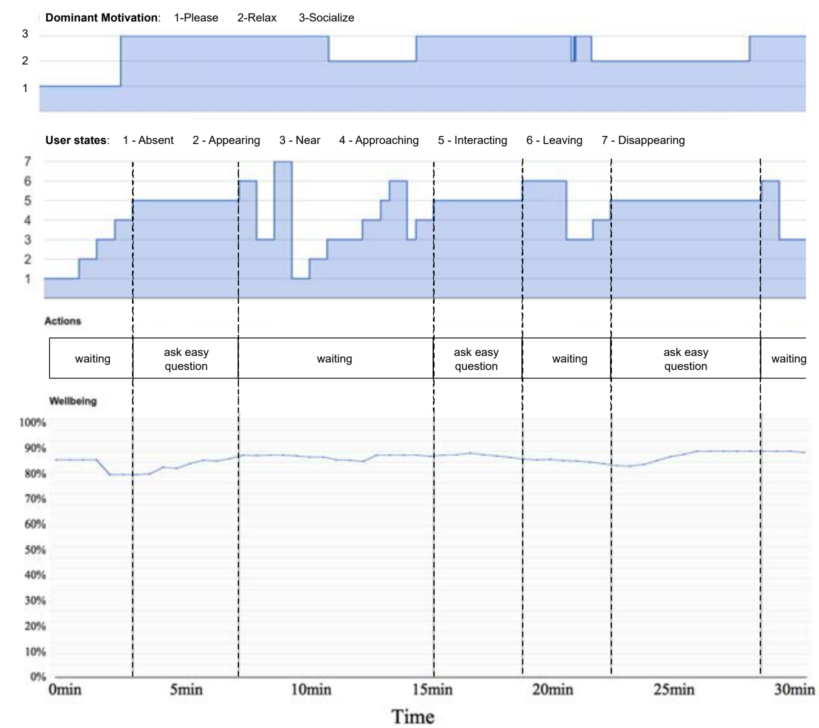


Figure 7: Detail of the exploitation phase with UP1.

motivation and the wellbeing decreases until the user starts interacting with Mini. Following the learned policy, the robot is waiting until it starts asking easy questions. As a consequence of the satisfaction of the user after the interactions, the dominant motivation changes to *socialize*. In this state, the robot keeps on asking easy questions until the user decides to leave. This behavior is repeated several times through the 30 minute session.

Around the minute 10, *relax* becomes the dominant motivation and Mini continues waiting even if it has the possibility of interaction sometimes.

As Mini learned, when *relax* is the dominant motivation but it has the possibility of interaction, it asks easy questions to UP1. This action outweighs others because it provides the largest increase in the robot's wellbeing.

In relation to the wellbeing, we have represented it as a percentage: the maximum value of 100% is the ideal situation when all drives are satiated. The minimum value of 0% corresponds to the worst situation when all drives have reached their saturation levels. Focusing on Mini's wellbeing (bottom plot on Figure 7), it is stabilized between 90% and 80%. This means that Mini has learned a policy that keeps its wellbeing in a very good range (recall that the ideal wellbeing is 100%).

Asking easy questions rises Mini's wellbeing due to its already mentioned two-sided effect on the drives *user's satisfaction* and *interaction*. For the action waiting, Mini executes it when *relax* is the internal state and the need of *rest* has to be reduced. However, since the other two drives can increase at a faster rate, the execution of this action can result on a reduction of the robot's wellbeing (first and third execution of waiting in Figure 7). However Mini has learned that this is the best action in this situation and this is corroborated by the stable high wellbeing Mini keeps through the exploitation session.

6. Conclusion

In this paper, we have proposed a bioinspired decision making system (DMS) that uses unsupervised learning to adapt the robot's behavior to different user's profiles and improve the HRI. In particular, we have considered a homeostatic process at the core of the DMS that includes the user's preferences in order to learn different policies of behavior for each user. The goal of the DMS is to maximize the robot's wellbeing, which is related to the drives: *user's satisfaction*, *rest*, and *interaction*. The DMS has been tuned to end up with a robot that is sociable and tries to please the multiple users.

Unexpected changes of the robot's state related to user's displacements (appearing, disappearing, approaching the robot, and moving away from the robot) have been modeled as time-based states. The robot has learned how to react to these exogenous actions.

The system has been tested with two different profiles of users: (i) UP1 approaches the robot sporadically and independently of the robot's action, and she likes to guess the right answer; and (ii) UP2 uses to approach the robot when Mini calls her attention and enjoys challenging questions. Mini has learned a different policy of action for each user that helps to keep its wellbeing within a high value as well as to enjoy both users, since their satisfaction is considered in the robot's wellbeing.

After the evaluation, we have observed that the number of times the robot explores the effects of the actions in all situations is a key aspect. We have observed that actions barely explored can lead to low performance (in terms of robot's wellbeing), and strange behaviors. This has happened, for instance, when the UP1 is disappearing and Mini is motivated to *please*; in this case the selected action is to attract the user's attention but it does not make sense since this user approaches, or moves away from, Mini almost randomly. We believe that a longer exploration of this state could lead to a different behavior.

In this line, we have observed that when the user behavior is not consistent, the robot learns a conservative policy of actions; that is, the most cost-effective actions are the preferred. This is the case of the action waiting in our scenario. It is worth mentioning the relevance of several parameters of the DMS; depending on how they are adjusted, the resulting robot's behavior could be different. In our case, we have ended up with a *lazy* robot that, in most of the states, is waiting but, at the same time, encourages users to interact with it.

6.1. Limitations

This work presents some limitations that constrain the results obtained.

The effects of the actions executed by the users, the exogenous actions, have been modeled as time-constrained transitional states. The time assigned to each one of these states is a design decision. Different times could affect the robot's behavior.

Moreover, the extension of the system to other exogenous actions is not straight forward. The consequences of each exogenous action have to be defined as new states.

Finally, the user profiles considered in this work are constant along all the phases, both learning and exploitation. We have not considered changes on the users' preferences or behaviors as it happens to humans in the course of their life.

Acknowledgment

The research leading to these results has received funding from the projects: Development of social robots to help seniors with cognitive impairment (ROBSEN), funded by the Ministerio de Economia y Competitividad (DPI2014-57684-R); and RoboCity2030-III-CM, funded by Comunidad de Madrid and cofunded by Structural Funds of the EU (S2013/MIT-2748).

References

[1] E. Cha, J. Forlizzi, S. S. Srinivasa, Robots in the home: Qualitative and quantitative insights into kitchen organization, in: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '15, ACM, New York, NY, USA, 2015, pp. 319–326. doi:10.1145/2696454.2696465.
URL http://doi.acm.org/10.1145/2696454.2696465

[2] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros, M. Vincze, Hobbit, a care robot supporting independent living at home: First prototype and lessons learned, Robotics and Autonomous Systems 75 (2016) 60 – 78, assistance and Service Robotics in a Human Environment. doi:https://doi.org/10.1016/j.robot.2014.09.029.
URL http://www.sciencedirect.com/science/article/pii/S0921889014002140

[3] P. Baxter, E. Ashurst, R. Read, J. Kennedy, T. Belpaeme, Robot education peers in a situated primary school study: Personalisation promotes child learning, PLOS ONE 12 (5) (2017) 1–23. doi:10.1371/journal.pone.0178126.
URL https://doi.org/10.1371/journal.pone.0178126

[4] F. Tanaka, A. Cicourel, J. R. Movellan, Socialization between toddlers and robots at an early childhood education center., in: Proceedings of the National Academy of Sciences of the United States of America, Vol. 104, 2007, pp. 17954–17958.

[5] A. Ogasawara, M. Gouko, Stationery holder robot that encourages office workers to tidy their desks, in: Proceedings of the 5th International Conference on Human Agent Interaction, HAI '17, ACM, New York, NY, USA, 2017, pp. 439–441. doi:10.1145/3125739.3132581.
URL http://doi.acm.org/10.1145/3125739.3132581

[6] A. R. Araujo, D. D. Caminhas, G. A. Pereira, An architecture for navigation of service robots in human-populated office-like environments, IFAC-PapersOnLine 48 (19) (2015) 189 – 194, 11th IFAC Symposium on Robot Control SYROCO 2015. doi:https://doi.org/10.1016/j.ifacol.2015.12.032.
URL http://www.sciencedirect.com/science/article/pii/S2405896315026567

[7] N. Mitsunaga, T. Miyashita, H. Ishiguro, K. Kogure, N. Hagita, Robovie-IV: A communication robot interacting with people daily in an office, IEEE International Conference on Intelligent Robots and Systems (2006) 5066–5072.

[8] S. Jeong, D. E. Logan, M. S. Goodwin, S. Graca, B. O'Connell, H. Goodenough, L. Anderson, N. Stenquist, K. Fitzpatrick, M. Zisook, L. Plummer, C. Breazeal, P. Weinstock, A social robot to mitigate stress, anxiety, and pain in hospital pediatric care, in: Proceedings of the Tenth An-

nual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, HRI'15 Extended Abstracts, ACM, New York, NY, USA, 2015, pp. 103–104. doi:10.1145/2701973.2702028. URL http://doi.acm.org/10.1145/2701973.2702028

[9] J. Messias, R. Ventura, P. Lima, J. Sequeira, P. Alvito, C. Marques, P. Carrio, A robotic platform for edutainment activities in a pediatric hospital, in: 2014 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 2014, pp. 193–198. doi:10.1109/ICARSC.2014.6849785.

[10] M. G. Kim, H. Lee, J. Lee, S. S. Kwak, Y. Joo, Effectiveness and service quality of robot museum through visitors experience: A case study of robolife museum in south korea, in: 2015 International Symposium on Micro-NanoMechatronics and Human Science (MHS), 2015, pp. 1–5. doi:10.1109/MHS.2015.7438289.

[11] R. Gehle, K. Pitsch, T. Dankert, S. Wrede, How to open an interaction between robot and museum visitor?: Strategies to establish a focused encounter in hri, in: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI '17, ACM, New York, NY, USA, 2017, pp. 187–195. doi:10.1145/2909824.3020219. URL http://doi.acm.org/10.1145/2909824.3020219

[12] C. Bartneck, J. Forlizzi, A design-centred framework for social human-robot interaction, in: RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), IEEE, 2004, pp. 591–594.

[13] A. V. Savkin, A. S. Matveev, M. Hoy, C. Wang, Safe Robot Navigation Among Moving and Steady Obstacles, Elsevier Inc., 2016.

[14] A. Saxena, J. Driemeyer, A. Y. Ng, Robotic grasping of novel objects using vision, The International Journal of Robotics Research 27 (2) (2008) 157–173.

[15] C. D. Frith, U. Frith, Implicit and Explicit Processes in Social Cognition, Neuron 60 (3) (2008) 503–510. doi:10.1016/j.neuron.2008.10.032.
URL http://linkinghub.elsevier.com/retrieve/pii/S0896627308009082

[16] M. H. Davis, Measuring individual differences in empathy: Evidence for a multidimensional approach., Journal of personality and social psychology 44 (1) (1983) 113.

[17] A. Gerace, A. Day, S. Casey, P. Mohr, An Exploratory Investigation of the Process of Perspective Taking in Interpersonal Situations, Journal of Relationships Research 4. doi:10.1017/jrr.2013.6.
URL http://www.journals.cambridge.org/abstract_S1838095613000061

[18] N. Eisenberg, P. A. Miller, The relation of empathy to prosocial and related behaviors., Psychological Bulletin 101 (1) (1987) 91–119. doi:10.1037/0033-2909.101.1.91.
URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.101.1.91

[19] M. P. Georgeff, F. F. Ingrand, Decision-Making in an Embedded Reasoning System, Proceedings of the 11th international joint conference on Artificial intelligence 2 (1989) 972–978.

[20] R. Brooks, Intelligence without representation, Artificial intelligence 47 (1991) 139–159.

[21] D. A. Norman, A. Ortony, D. M. Russell, Affect and machine design: Lessons for the development of autonomous machines, IBM Systems Journal 42 (1) (2003) 38–44.

[22] A. Ortony, D. A. Norman, W. Revelle, Affect and Proto-Affect in Effective Functioning, in: Who Needs Emotions?, Oxford University Press, 2005, pp. 173–202.

[23] A. Sloman, M. Scheutz, B. Logan, Evolvable Architectures For Human-Like Minds, Affective minds (2000) 169–181.

[24] K. C. Berridge, Motivation concepts in behavioral neuroscience, Physiology and Behavior 81 (2) (2004) 179–209.

[25] W. Cannon, The wisdom of the body, W. W. Norton and Company, 1932.

[26] J. D. Velásquez, Modeling Emotions and Other Motivations in Synthetic Agents, Fourteenth National Conference on Artificial Intelligence (1997) 10.

[27] R. C. Arkin, K. Ali, A. Weitzenfeld, F. Cervantes-Perez, Behavioral models of the praying mantis as a basis for robotic behavior, Robotics and Autonomous Systems 32 (1) (2000) 39 – 60. doi:https://doi.org/10.1016/S0921-8890(99)00121-9.
URL http://www.sciencedirect.com/science/article/pii/S0921889099001219

[28] R. C. Arkin, M. Fujita, T. Tagaki, R. Hasegawa, An Ethological and Emotional Basis for Human- Robot Interaction, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002), Vol. 42, 2002, pp. 191–201.

[29] A. Stoytchev, R. C. Arkin, Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280, U.S.A., Incorporating Motivation in a Hybrid Robot Architecture, Journal of Advanced Computational Intelligence and Intelligent Informatics 8 (3) (2004) 269–274. doi:10.20965/jaciii.2004.p0269.
URL https://www.fujipress.jp/jaciii/jc/jacii000800030269

[30] D. Cañamero, Designing Emotions for Activity Selection, Emotions in humans and artifacts (2003) 115–148.

[31] D. Cañamero, Modeling motivations and emotions as a basis for intelligent behavior, in: Proceedings of the first international conference on Au-

tonomous agents - AGENTS '97, no. Abbott 1884, ACM Press, New York, New York, USA, 1997, pp. 148–155.

[32] D. Canamero, A hormonal model of emotions for behavior control, VUB AI-Lab Memo 2006 (1997) 1–10.

[33] C. L. Breazeal, Designing sociable robots, MIT Press, 2004.

[34] Internal robotics, Connection Science 16 (4).

[35] V. Vouloutsi, S. Lallée, P. F. M. J. Verschure, Modulating behaviors using allostatic control, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8064 LNAI (2013) 287–298.

[36] H.-L. Cao, P. Gómez Esteban, D. B. Albert, R. Simut, G. Van de Perre, D. Lefeber, B. Vanderborght, A Collaborative Homeostatic-Based Behavior Controller for Social Robots in HumanRobot Interaction Experiments, International Journal of Social Robotics 9 (5) (2017) 675–690. doi:10.1007/s12369-017-0405-z.
URL https://doi.org/10.1007/s12369-017-0405-z

[37] F. Gomez-Donoso, S. Orts-Escolano, A. Garcia-Garcia, J. Garcia-Rodriguez, J. A. Castro-Vargas, S. Ovidiu-Oprea, M. Cazorla, A robotic platform for customized and interactive rehabilitation of persons with disabilities, Pattern Recognition Letters 99 (2017) 105 – 113, user Profiling and Behavior Adaptation for Human-Robot Interaction. doi:https://doi.org/10.1016/j.patrec.2017.05.027.
URL http://www.sciencedirect.com/science/article/pii/S0167865517301903

[38] C. Hieida, T. Horii, T. Nagai, Decision-Making in Emotion Model, in: Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, ACM, 2018, pp. 127–128.

[39] M. Heerink, B. Kröse, V. Evers, B. Wielinga, Assessing acceptance of assistive social agent technology by older adults: the almere model, International Journal of Social Robotics 2 (4) (2010) 361–375. doi:10.1007/s12369-010-0068-5.
URL https://doi.org/10.1007/s12369-010-0068-5

[40] M. Heerink, How elderly users of a socially interactive robot experience adaptiveness, adaptability and user control, in: 2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI), 2011, pp. 79–84. doi:10.1109/CINTI.2011.6108476.

[41] S. Rossi, F. Ferland, A. Tapus, User profiling and behavioral adaptation for hri: A survey, Pattern Recognition Letters 99 (2017) 3 – 12, user Profiling and Behavior Adaptation for Human-Robot Interaction. doi:https://doi.org/10.1016/j.patrec.2017.06.002.
URL http://www.sciencedirect.com/science/article/pii/S0167865517301976

[42] A. Vinciarelli, A. S. Pentland, New social signals in a new interaction world: The next frontier for social signal processing, IEEE Systems, Man, and Cybernetics Magazine 1 (2) (2015) 10–17. doi:10.1109/MSMC.2015.2441992.

[43] A. Coninx, P. Baxter, E. Oleari, S. Bellini, B. Bierman, O. Blanson Henkemans, L. Caamero, P. Cosi, V. Enescu, R. Ros Espinoza, A. Hiolle, R. Humbert, B. Kiefer, I. Kruijff-Korbayov, R. Looije, M. Mosconi, M. Neerincx, G. Paci, G. Patsis, C. Pozzi, F. Sacchitelli, H. Sahli, A. Sanna, G. Sommavilla, F. Tesser, Y. Demiris, T. Belpaeme, Towards Long-Term Social Child-Robot Interaction: Using Multi-Activity Switching to Engage Young Users, Journal of Human-Robot Interaction 5 (1) (2015) 32. doi:10.5898/JHRI.5.1.Coninx.
URL http://dl.acm.org/citation.cfm?id=3109941

[44] D. Bacciu, C. Gallicchio, A. Micheli, M. D. Rocco, A. Saffiotti, Learning context-aware mobile robot navigation in home environments, in:

IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications, 2014, pp. 57–62. doi:10.1109/IISA.2014.6878733.

[45] A. Tanevska, F. Rea, G. Sandini, A. Sciutti, Towards an Affective Cognitive Architecture for Human-Robot Interaction for the iCub Robot, in: 1st Workshop on Behavior, Emotion and Representation: Building Blocks of Interaction, 2017.

[46] A. Ramachandran, N. Haven, B. Scassellati, Adapting Difficulty Levels in Personalized Robot-Child Tutoring Interactions, Machine Learning for Interactive Systems (2014) 56–59.

[47] O. Grr, B. Rosman, F. Sivrikaya, S. Albayrak, Social Cobots: Anticipatory Decision-Making for Collaborative Robots Incorporating Unexpected Human Behaviors, in: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, ACM, 2018, pp. 398–406.

[48] Á. Castro-González, M. Malfaz, M. A. Salichs, Learning the Selection of Actions for an Autonomous Social Robot by Reinforcement Learning Based on Motivations, International Journal of Social Robotics 3 (4) (2011) 427–441.

[49] Á. Castro-González, M. Malfaz, M. Á. Salichs, An autonomous social robot in fear, IEEE Transactions on Autonomous Mental Development 5 (2) (2013) 135–151.

[50] A. Castro-Gonzalez, M. Malfaz, J. F. Gorostiza, M. A. Salichs, Learning behaviors by an autonomous social robot with motivations, Cybernetics and Systems 45 (7) (2014) 568–598. arXiv:https://doi.org/10.1080/01969722.2014.945321, doi:10.1080/01969722.2014.945321.
URL https://doi.org/10.1080/01969722.2014.945321

[51] C. L. Hull, Principles of Behavior: An Introduction to Behavior Theory, Vol. 25 of The Century psychology series, Appleton-Century, 1943.

[52] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, Press, MIT, 2012.

[53] S. C. Gadanho, P. Dayan, Learning behavior-selection by emotions and cognition in a multi-goal robot task, Journal of Machine Learning Research 4 (2003) 385–412.

[54] R. Pérula-Martínez, E. Salichs, I. P. Encinar, Á. Castro-González, M. A. Salichs, Improving the Expressiveness of a Social Robot through Luminous Devices, in: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts - HRI'15 Extended Abstracts, ACM Press, 2015, pp. 5–6.