Mathematical Modelling of Theory of Mind

Enabling Socially Assistive Robots to understand and predict humans in long-term interactions

Maria Luís Morão Patrício

TUDelft

Technische Universiteit Delft

Mathematical Modelling of Theory of Mind

Enabling Socially Assistive Robots to understand and predict humans in long-term interactions

by

Maria Luís Morão Patrício

Student Number: 5143640
to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday December 14, 2021 at 13:00 PM.

Student number:	5143640	
Project duration:	March 8, 2021 – December 14, 2021	
Thesis committee:	Prof. dr. ir. M. Mulder,	TU Delft, chair
	Dr. A. Jamshidnejad,	TU Delft, supervisor
	Prof. dr. ir. B. De Schutter,	TU Delft, external examiner
	Dr. A. Dabiri,	TU Delft

Cover Image: https://listimg.pinclipart.com/picdir/s/520-5208843_brain-learning-png-clipart-artificial-intelligence-logo-png.png and https://www.pngitem.com/middle/Tombwbw_it-s-in-our-pipeline-to-roll-out/

Preface

Socially assistive robots i.e., robots that assist humans via social interactions, have shown great potential to help us in fields such as education, healthcare and in our daily lives. However, they are still not able to assist us in the real world, given that they struggle to act socially and humanly. The present project proposes to tackle this issue by formalising, formulating and implementing mathematical models of human cognition and behaviour. These models can be applied in socially assistive robots to allow them to properly understand humans in social interactions and, consequently, behave naturally in these interactions.

This thesis marks the end of my MSc in Aerospace Engineering. After these two years, I can say that it was a great challenge to move to another country, start a degree in a new university, make new friends, adapt to a new culture and finish my degree during a global pandemic. However, I believe that all these challenges contributed to make me grow both on a personal and on an academic level. This project exceeded all my expectations: I imagined I would learn how to carry out a research project, but I ended up learning so much more. I would like to thank my supervisor, Anahita Jamshidnejad, not only for guiding me through this project, but also for going the extra mile to make me understand what it means to be a researcher. I want to thank Riccardo for the constant support during these two years, for believing in me even when I did not, and for always reminding me that everything will be fine. I want to thank João for being one of my first friends in Delft, for the guidance in the first weeks, and for all the other hangouts since then. Thank you to Matteo for bringing some fun to the study sessions, for being my internship buddy, and for always providing road assistance every time my bike broke down. I also want to thank all the participants that took part in the online survey, which provided the data that allowed the validation of the models. I want to thank the friends I made through my BSc, for making the three years that preceded the MSc, which were crucial to define my path, so enjoyable. I would also like to thank my childhood friends, for always keeping in touch and never allowing me to forget where I come from. Special thanks to Henrique, for helping me with designing the cover image. Finally, I owe a very special thank you to my family, specially to my parents, Elsa and Luís, and my siblings, Leonor, Henrique and Miguel, for always supporting me in chasing my ambitions, even if this meant that I would be living so far away from them. I would not have managed to be where I am without their support, and I am eternally thankful for that.

Maria Luís Morão Patrício
Delft, November 2021

Contents

List of Acronyms

AI Artificial Intelligence

ASD Autism Spectrum Disorder

BN Bayesian Network

DBN Dynamic Bayesian Network

ToM Theory of Mind

BToM Bayesian Theory of Mind

DAG Direct Acyclic Graph

HRI Human-Robot Interaction

MDP Markov Decision Process

POMDP Partially Observable Markov Decision Process

HPOMDP Hierarchical Partially Observable Markov Decision Process

FLC Fuzzy Logic Control

RL Reinforcement Learning

AR Assistive Robot

SAR Socially Assistive Robot

SIR Socially Interactive Robot

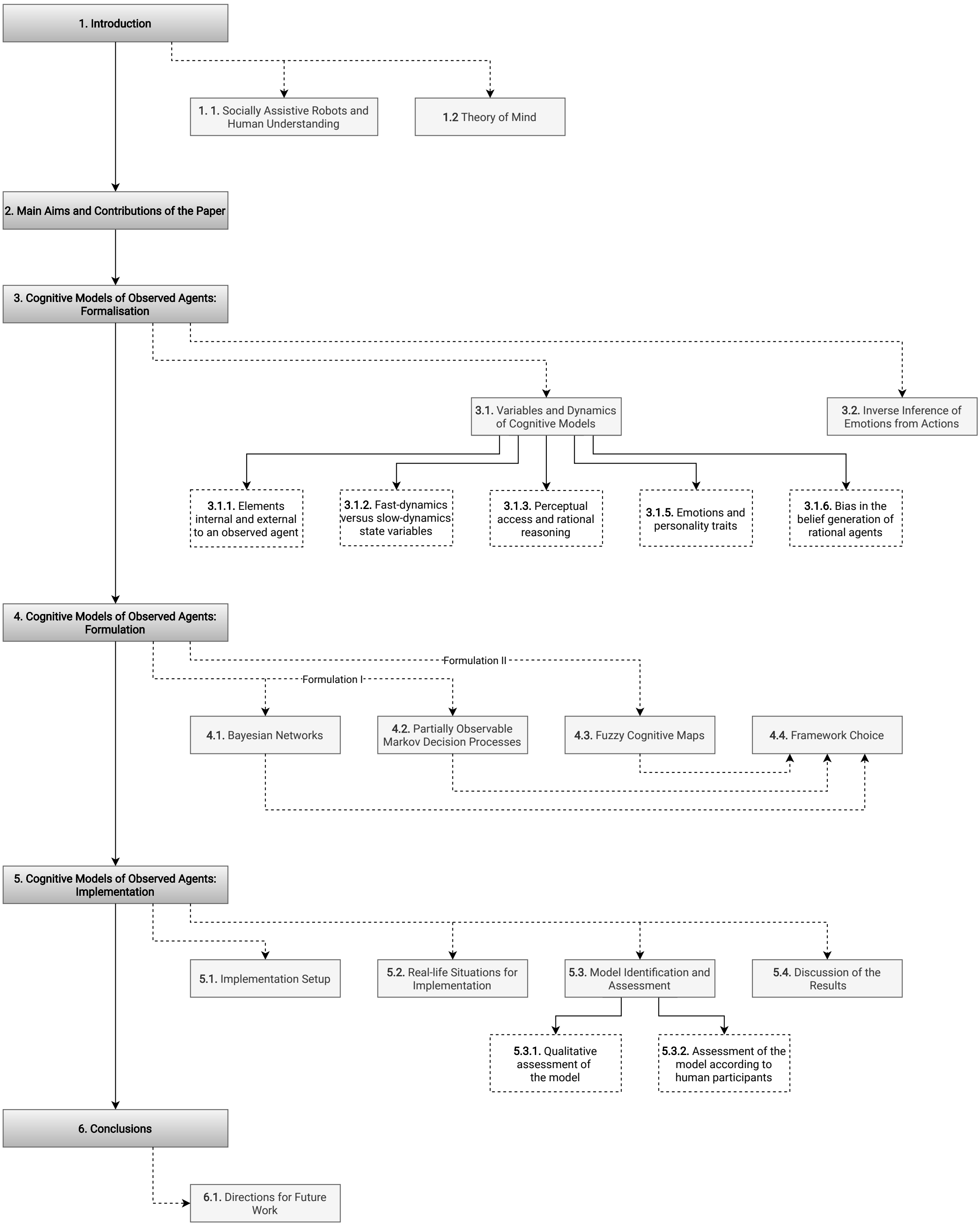
I

Thesis Article

1

Article Road Map

Article Road Map


1. Introduction
1. 1. Socially Assistive Robots and Human Understanding
1.2 Theory of Mind
2. Main Aims and Contributions of the Paper
3. Cognitive Models of Observed Agents: Formalisation
3.1. Variables and Dynamics of Cognitive Models
3.2. Inverse Inference of Emotions from Actions
3.1.1. Elements internal and external to an observed agent
3.1.2. Fast-dynamics versus slow-dynamics state variables
3.1.3. Perceptual access and rational reasoning
3.1.5. Emotions and personality traits
3.1.6. Bias in the belief generation of rational agents
4. Cognitive Models of Observed Agents: Formulation
Formulation II
Formulation I
4.1. Bayesian Networks
4.2. Partially Observable Markov Decision Processes
4.3. Fuzzy Cognitive Maps
4.4. Framework Choice
5. Cognitive Models of Observed Agents: Implementation
5.1. Implementation Setup
5.2. Real-life Situations for Implementation
5.3. Model Identification and Assessment
5.4. Discussion of the Results
5.3.1. Qualitative assessment of the model
5.3.2. Assessment of the model according to human participants
6. Conclusions
6.1. Directions for Future Work

2

Mathematical Modelling of Theory of Mind: Enabling Socially Assistive Robots to understand and predict humans in long-term interactions

Mathematical Modelling of Theory of Mind

Enabling socially assistive robots to understand and predict humans in long-term interactions

Maria L. Morão Patrício

Abstract Socially Assistive Robots (SARs) i.e., robots that assist humans through social interactions with them, have shown potential to improve the quality of life of their users. Nonetheless, the state-of-the-art SARs face challenges that prevent them from or limit them in assisting humans in the real world. Although current SARs display cues that simulate social behaviours, the interactions that they generate are often not realistic and fail to engage their users for long periods of time. Most of the challenges currently faced by SARs are a consequence of the lack of social awareness and understanding of the cognitive procedures of individual users. In human interactions, people develop cognitive models of each other to achieve social goals and to respond to the needs of their peers according to their emotional states. Therefore, the present research project proposes to develop a mathematical model of human cognition, based on Theory of Mind (ToM), and to implement it for SARs. To do so, the cognitive processes behind human behaviour in social interactions are carefully investigated and formalised, and a model of human cognition is presented. Two formulations of the model are proposed and compared. The model is implemented using Fuzzy Cognitive Map (FCM), and is then qualitatively analysed and quantitatively assessed. It is concluded that, to accurately estimate the cognitive states of humans, the model must comprise user-specific variables that describe agents throughout interactions, the linkages of the FCM must be represented as functions, and the model must be personalised to every user.

Keywords Theory of Mind · Socially Assistive Robots · Long-term Interactions · User Modelling · Cognitive Model

1 Introduction

1.1 Socially Assistive Robots and Decision-Making

Socially Assistive Robots (SARs) are a subclass of robots that assist human users via social interactions (Feil-Seifer and Matarić 2005). SARs provide help to users in fields such as health care (e.g., rehabilitation and therapy (Tapus and Mataric 2008)), education (e.g., tutoring and therapy for children with cognitive impairments (Clabaugh et al. 2019; Scassellati et al. 2018)) and daily life assistance (e.g., in home-assistance and weight loss programs (Kidd and Breazeal 2008)). The motivation behind the usage of this approach is that, in the fields of application of SARs, engaging in social interactions with the users boosts the therapeutic and learning outcome of the sessions (Clabaugh et al. 2019; Tapus and Mataric 2008).

The majority of the studies about SARs focuses on short-term interactions (i.e., interactions composed by few sessions, usually shorter than a month), since they are simpler to carry out and consume little time (Leite et al. 2013). Nonetheless, in education and health care, users often require support for a longer period of time so that meaningful therapeutic or learning results are achieved. Consequently, studies have been carried out to develop and analyse long-term interactions between humans and SARs (Kidd and Breazeal 2008; Scassellati et al. 2018). These long-term studies have identified a major setback in long-term human-SAR interactions: SARs cannot maintain the users engaged for periods of time longer than a month (Leite et al. 2013; Clabaugh et al. 2019). In education and health recovery, the processes in which SARs are meant to assist the users with are usually longer than the periods of time for which these robots can engage the attention of the users. Consequently, the assistance provided by these robots is not as effective as it could be if the SARs were able to engage the attention of the users indefinitely (Leite et al. 2013; Scassellati et al. 2018; Kidd and Breazeal 2008).

The lack of capability to maintain the user engaged is a consequence of the simple and rudimentary nature of the social skills that are displayed by these robots. Some projects have aimed to develop SARs that act as social agents by displaying behaviours and non-verbal cues that are often seen in human interactions, such as joint-attention (Scassellati et al. 2018), eye contact (Kidd and Breazeal 2008) and facial expressions (Clabaugh et al. 2019). Nonetheless, these approaches fail to recognise the ideal moments to show such behaviours and cues, since they do not focus on understanding the users (Leite et al. 2013). Having recognised that every user may require a different rehabilitation or teaching approach, some projects have aimed to personalise the controllers of the robots to every user (Tapus and Mataric 2008; Clabaugh et al. 2019; Scassellati et al. 2018). However, the approaches that were used in these research projects were simply based on maximising the task performance rather than on analysing the user behaviour and characteristics. In general, socially assistive robotics designs controllers that are tailored to the specific task in which the robot must assist the human. Therefore, they lack the capability to act naturally and socially in general, quotidian contexts.

In order to be capable of truly acting as social agents, SARs must be able to understand the mechanics behind human interactions (Matarić and Scassellati 2016), as well as being aware of the affective states of the users (Leite et al. 2013). So far, the control approaches that are used to steer the behaviour of the SARs are mostly model-free approaches, such as Reinforcement Learning (Tapus and Mataric 2008; Clabaugh et al. 2019) and rule-based approaches (Scassellati et al. 2018; Kidd and Breazeal 2008). Socially assistive robotics is attempting to develop robots that act as humans and as social agents, without implementing a tool that humans have: an extensive understanding of how the cognitive processes and mental states of their peers shape their actions and behaviours. Consequently, socially assistive robotics needs a thorough model of the cognitive processes that drive human behaviour. The present research project focuses on developing such a model.

1.2 Theory of Mind

Interactions among humans are mainly based on empathising and mutual understanding of one another, i.e., being able to make predictions about other person's upcoming actions and state-of-mind, as well as inverse inference of the other person's current state-of-mind based on their observable actions. The theory that adheres to this explanation is known as the Theory of Mind (ToM) (Scassellati 2002). Prior studies agree that ToM can be described based on the *principle of rational action*, which states that humans act *approximately* rationally towards maximising the fulfilment of their goals, given their beliefs about the external world (Dennett 1987; Baker et al. 2011; Jara-Ettinger et al. 2016; Saxe and Houlihan 2017). Principle of rational action is an idealised conception of human's behaviour. In this paper, a cognitive agent that is assumed to behave based on the principle of rational action is called a *rational agent.*

Based on ToM and the assumption that the beliefs and goals of a rational agent result from a combination of external conditions and the perceptual access that the rational agent has to those conditions (*principle of rational belief*), Baker et al. (2011) proposed and implemented a Bayesian ToM model resorting to the concept of Partially Observable Markov Decision Process (POMDP). By inverting the POMDP using Bayesian Inference, the developed model managed to infer jointly the beliefs and goals of rational agents moving in two-dimensional spaces based on their actions. Furthermore, the estimated beliefs and goals were characterised by a prediction of the agent's certainty about the belief and its level of desire of attaining the goal, respectively. In various experiments, the inferences of the goals and beliefs made by the aforementioned model - based on Bayesian ToM - showed good similarities with the inferences made by humans, which acknowledges that goals and beliefs are interdependent and should, therefore, be inferred simultaneously. Additionally, Baker (2012) showed via further experiments that the same model can be used to predict the actions of a rational agent based on its inferred beliefs and goals. Nonetheless, these investigations fall short to address the implementation of such a model in a realistic scenario that

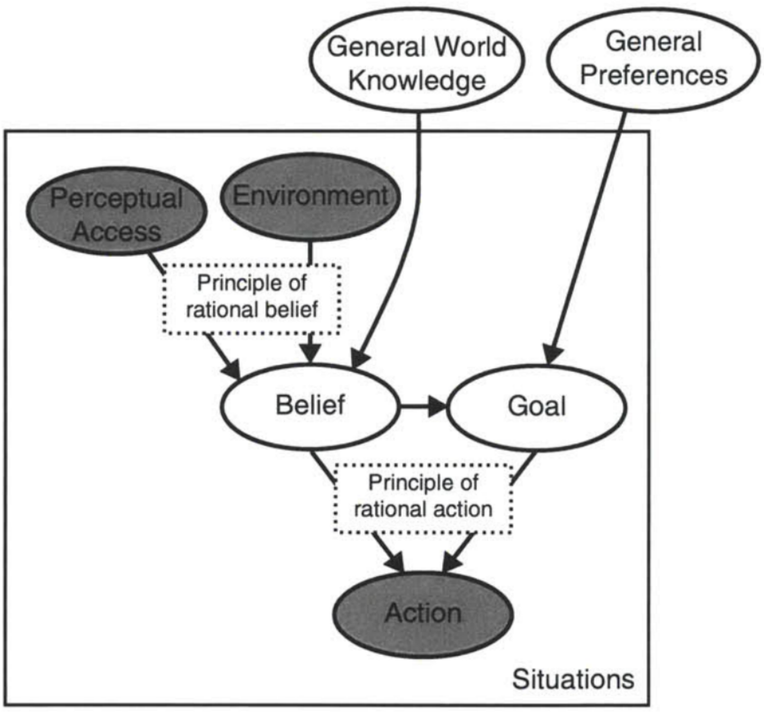


Fig. 1: A cognitive model for a rational agent: This model explains the behaviour of a rational agent based on the principle of rational action (i.e., intentional actions of the agent are assumed to be influenced by the agent's beliefs and goals, where the agent aims at optimising the fulfilment of its goals upon its actions, given its beliefs), the principle of rational belief (i.e., beliefs and goals of a rational agent result from both the surrounding environment and the perceptual access the agent has to its environment), and the influence of general world knowledge and general preferences on, respectively, beliefs and goals of the rational agent (Baker 2012).

simulates representative human interactions. Despite the simple environment in which the experiments took place, the performance shown by the model indicates a strong potential of broadening this formulation of the Bayesian ToM to model the cognition and behaviour of rational agents, including humans, in real-life situations.

As an extension to the previous model, Baker (2012) proposed a model based on the network represented in Figure 1, in which the beliefs and goals of a rational agent are additionally influenced by, respectively, *general world knowledge* and *general preferences* of the agent. The general world knowledge and general preferences of a rational agent are assumed to be "permanent" or "fixed" for a rational agent, i.e., they do not vary across different situations. Baker (2012) refers to these as high-level variables, since they are constant over time for each agent but distinct for different agents. On the contrary, the beliefs and goals of the rational agent may regularly change as a response to various situations. Although the influence of these two high-level variables was recognised by Baker (2012), the variables were not included in the model. When interacting with an agent for a short period of time, it is difficult to accurately estimate the aforementioned high-level variables, since understanding an agent's personality requires interacting with that agent for longer. Hence, when working with short-term interactions, which is the case of the study carried out by Baker (2012), the inclusion of the high-level variables would increase the complexity of the model without providing any benefit. Therefore, only the variables inside the *situation box* (see Figure 1) were considered in their experiments.

Nonetheless, given the importance of personalisation in long-term interactions between rational agents including human beings (Leite et al. 2013), developing an extended version of the discussed model - in order to encompass various parameters/variables that are inherent to a rational agent and, consequently, are permanent across interactions and can play an important role in the individual cognitive procedure of every rational agent - is of utmost importance. As mentioned earlier, ToM is composed of two main procedures: (1) inferring the state-of-mind of a rational-agent based on the observed behaviour of that agent, called *inverse inference*; and (2) predicting the upcoming actions of a rational agent based on the current (inferred) state-of-mind of the agent, called *forward inference* (Saxe and Houlihan 2017).

This paper describes the processes of formalising, formulating and implementing a mathematical model of human cognition that can be applied in socially assistive robotics. Section 2 describes the main contributions of the present paper to the relevant research fields. Section 3 describes the formalisation process of the mathematical cognitive models and is divided in two parts: while Section 3.1 analyses the variables and dynamics of the cognitive models, as well as the relationship between these variables, Section 3.2 defines a secondary module that can be used in parallel with the main model to improve the accuracy of its predictions. Section 4 analyses two possible formulations of the suggested mathematical cognitive models and explains the choice of the formulation used to implement the models in this paper. Section 5 describes the implementation and assessment of the models, and discusses the results obtained. Finally, the conclusions and directions for future work are presented in Section 6. A more detailed outline of the paper can be found in the previous chapter Article Road Map.

2 Main Aims and Contributions of the Paper

Our main focus is on the development of mathematical cognitive models that represent human-like interactions for rational agents. In this paper, a rational agent that infers about the state-of-mind and actions/behaviours of another rational agent is called an *observer agent*, and the rational agent whose state-of-mind and actions/behaviours are being inferred is called an *observed agent*. One application of such mathematical cognitive models is for interactions between a socially assistive robot and a human, where the robot (observer agent) infers about the human (observed agent) according to such models. The main contributions of this paper include:

- Formalising the first comprehensive mathematical models that take into account the deep cognitive processes that define the behaviour of humans, and that can be used to estimate the mental states of rational agents, as well as to predict their behaviour.
- Defining a framework that allows the models to be represented using state space modelling and, consequently, to be implemented with a wide range of control approaches, such as model predictive control, to command the behaviour of a SAR.
- Implementing the mathematical cognitive models using Fuzzy Cognitive Maps (FCMs) and assessing the accuracy of their predictions regarding three cognitive variables of human participants.

The resulting cognitive models contribute to the research on ToM, and to developing analyses and model-based control approaches that yield long-term engaging human-machine interactions and more realistic, human-like machine-machine interactions.

3 Cognitive Models of Observed Agents: Formalisation

Human-like rational cognitive procedures should estimate the current state-of-mind of an observed agent, as well as predict its future state-of-mind and actions/behaviours. Fixed (or more accurately, valid in longer terms) characteristics (e.g., personality traits, general preferences, and general information and knowledge available for the agent) and dynamic (i.e., varying more frequently in time) inner variables (e.g., goals, beliefs, and emotions) of the observed agent, as well as their interdependencies, influence the agent's state-of-mind and thus actions/behaviours. Hence, they should both be incorporated into a cognitive model. Moreover, such a model should represent the effect of environmental data that is perceived in a personalised way by the observed agent.

3.1 Variables and Dynamics of Cognitive Models

In order to provide a comprehensive and accurate formalisation of human-like rational cognition, we discuss and analyse a number of representative examples. Accordingly, we develop a network representation composed of elements that are connected via directed links (which represent the interdependencies and influences of these elements). The resulting network representation will later on be used to mathematically formulate the cognitive model based on various mathematical theories, in particular, probability theory and fuzzy logic theory. The elements of the network correspond to:

1. External (uncontrolled) inputs of the cognitive model: Examples include environmental factors, e.g., weather condition, that may influence the state-of-mind and thus actions/behaviours of the observed agent.
2. State variables of the cognitive model: Examples include the state-of-mind of the observed agent, e.g., beliefs, goals, and emotions.
3. Fixed parameters (or more accurately, variables of low-frequency dynamic) of the cognitive model: Examples include personality traits, general preferences, and general world knowledge of the observed agent.
4. Dynamic processes: These are functions that receive the current external inputs, state variables, and fixed parameters of the cognitive model and update the next-step state variables of the observed agent or predict its upcoming actions/behaviours. These functions can be formulated based on the principle of rational action and the principle of rational belief (see Baker et al. (2011)).

3.1.1 Elements internal and external to an observed agent

Baker (2012) defines the elements of the cognitive network representation given in Figure 1 based on whether they are dependent on a specific situation or not. We, instead, differentiate the elements of the cognitive model considering whether they are external or internal to the observed agent (see Figure 2). The main advantages of our proposed approach include:

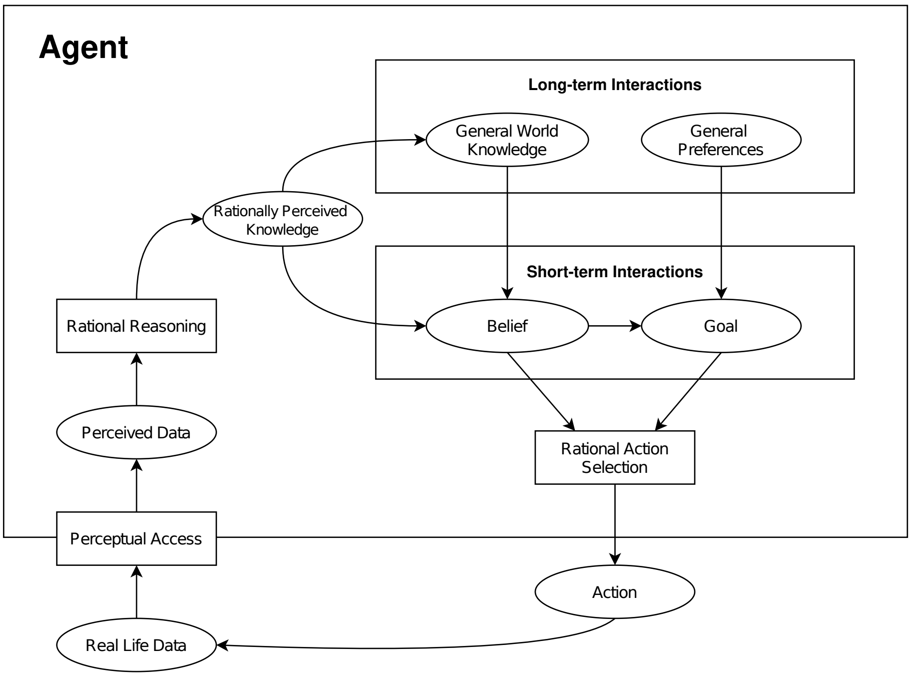


Fig. 2: Proposed cognitive model incorporating input and output variables, unobservable state variables, *personalised* perceptual access, rational reasoning and rational action selection, and their interdependencies and connections. Variables and processes are represented by, respectively, ovals and rectangles. Internal elements (unobservable for an observer agent), external elements (observable for an observer agent), and partially external elements (partially observable for an observer agent) are illustrated inside, outside, and on the border of the *agent* box.

- This categorisation allows developing mathematical cognitive models that correspond to existing frameworks of mathematical modelling and systems theory, e.g., state space modelling.
- The internal elements of the observed agent (i.e., goals, beliefs, general world knowledge, general preferences) are not observable by the observer agent. Inference of these unobservable elements is in general personalised and unique to an observer agent. Although we do not consider the personalisation of the observable agent, such personalisation can be useful when the consideration of second order inferences is important i.e., inferences about the inferences made by an observer agent about an observed agent (Baker et al. 2008). The proposed categorisation allows to systematically incorporate this personalisation into the modelling.
- The external elements (e.g., the environment and actions of the observed agent) are observable by the observer agent. Similarly, this categorisation can be important when the personalisation of the model to the observed agent is relevant (e.g., second-order inferences), given that the elements considered observable can be assumed to be equal for all the observed agents.
- Baker (2012) develops a cognitive model (see Figure 1) based on the principles of rational belief and rational action for the observed agent, i.e., assuming *universal* and *identical* rationalisation for different observed agents. Our categorisation allows considering personalised rationalisation and action selection [1] processes as internal elements for observed agents, resulting in more realistic cognitive models.

In addition to fully internal (i.e., unobservable) and fully external (i.e., observable) elements, our representation can incorporate partially external (i.e., partially observable) elements. More specifically, the perceptual access of an observed agent is partially external, given that this process is influenced by external inputs, and is partially internal, since it is shaped by internal characteristics of the individual. This dual influence on perceptual access is explained in detail in Section 3.1.3.

3.1.2 Fast-dynamics versus slow-dynamics state variables

The state variables of the proposed cognitive model are distinguished according to their relevance for different (i.e., short-term or long-term) interactions between two rational agents and to the frequency of their dynamics. Consequently, two categories of state variables are defined:

1. Fast-dynamics state variables, which may constantly vary (with a time scale in the range of seconds or minutes) as a response to specific situations the observed agent faces. Goals and beliefs (also considered by Baker (2012)) and emotions (considered in this paper, but not by Baker (2012)) are fast-dynamics state variables.
2. Slow-dynamics state variables, which vary according to large time scales (months or years). General preferences and general world knowledge (also proposed by Baker (2012)), as well as personality traits (considered in this paper, but not by Baker (2012)) are slow-dynamics state variables.

Remark 1 Slow-dynamics state variables may influence the evolution of fast-dynamics state variables, while the opposite is not necessarily true (especially in short terms). Note that the aim of this paper is to formalise and formulate the dynamic evolution of fast-dynamics state variables, and that modelling the evolution of slow-dynamics state variables is out of the scope of this paper. Thus, slow-dynamics state variables are mainly considered as fixed parameters in the proposed cognitive models.

Remark 2 When an observer agent only interacts with an observed agent for a short period of time, slow-dynamics state variables are not as relevant to understand the observed agent as fast-dynamics state variables. Thus, fast-dynamics state variables are more relevant for short-term interactions. Contrarily, slow-dynamics state variables are more relevant to know and understand an observed agent throughout long-term interactions than fast-dynamics state variables, since the patterns in the behaviour of the observed agent are more important than the transient state of mind of that agent in every circumstance.

[1] We use the term *rational action selection* rather than *principle of rational action*, as we do not consider it a universal principle.

For the variables introduced by Baker (2012), goals and beliefs are considered fast-dynamics and general preferences and general world knowledge are considered slow-dynamics state variables in this paper. Goals refer to immediate desires and needs of rational agents, such as finding food, grabbing an object, meeting someone, reaching a location. General preferences of a rational agent form in long terms and remain invariant for longer. Moreover, to identify general preferences, several interactions with the agent are needed. Examples of general preferences include favourite tastes, people whom a rational agent likes or dislikes, routines, habits, and hobbies of the rational agent. Beliefs correspond to *temporary* interpretation or knowledge gained by a rational agent from its world, while general world knowledge consists in *persistent* rationally perceived knowledge of the rational agent, which is fixed or rarely updated. For example, while the current location of a friend who has left an hour ago to go to the drugstore is a belief, the location of the drugstore is general world knowledge. Note that goals may depend on the beliefs of a rational agent, while they may be influenced by the agent's general preferences as well.

In this paper, two additional state variables are considered for an observed agent: emotions, which are fast-dynamics, and personality traits, which are slow-dynamics. Moreover, a main contribution of this paper is incorporating the influence of the slow-rate state variables in evolution of the dynamics of the fast-rate state variables. Therefore, before discussing the new state variables, we discuss the following examples that show the importance of considering these slow-dynamics state variables for modelling the cognitive procedures and reasoning of a rational agent.

In these examples, the observer agent (described by first-person pronouns) infers about the goal of the observed agent (Ana) based on the observed actions (*inverse inference*).

Example 1 Ana and I are both in the library at 5:00 PM. Ana picks up her wallet and walks towards the door (*action of the observed agent noticed by the observer agent*). The coffee house nearby has a late opening hour until 6:00 PM, but I do not know whether Ana knows this (no access to general world knowledge of the observed agent). I guess Ana knows about the opening hour of the coffee house and believes that it is still open (*guessing the general world knowledge and inferring about the belief of the observed agent*). I infer that Ana is going to buy a cup of coffee (*goal of the observed agent inferred by the observer agent*).

In this example, since the observer agent does not have access to the general world knowledge of the observed agent, the inference involves an intermediate procedure, i.e., inference about the belief of the observed agent, which is based on a guess - rather than facts - that the observer agent makes about the general world knowledge of the observed agent.

Example 2 Consider Example 1, but this time Ana has told me before that she knows the coffee house nearby is open until late (*access to the general world knowledge of the observed agent*). Based on Ana's observed actions and general world knowledge, I suppose with a higher certainty that Ana assumes that the coffee house is open (*belief of the observed agent deduced by the observer agent*) and I infer with a higher certainty that she is going to buy a cup of coffee.

This example shows that when the general world knowledge of the observed agent is known by the observer agent, the inverse inference about the goals of the observed agent are less prone to uncertainties. With the next example, however, we show that knowing the general world knowledge of the observed agent alone may not be sufficient to make an accurate inverse inference.

Example 3 In Example 2, if in addition to being aware of Ana's general world knowledge, I know she likes to drink coffee while studying (*access to the general preference of the observed agent*), then my inference about Ana planning to buy coffee is prone to even lower uncertainty than in Example 2. However, if I have seen before that Ana never drinks coffee (*general preference of the observed agent*) making inferences related to the coffee house for her beliefs and consequently goals is unlikely. While my additional information about Ana's general preference in the first case supported the certainty of my inference about her goals, this time it prevents me from making an erroneous inference about her goals.

The last example shows that access of the observer agent to the slow-dynamics state variables (i.e., general world knowledge and general preferences) of the observed agent significantly improves the reliability and level of certainty of the inferred fast-dynamics state variables (i.e., beliefs and goals). Moreover, having access to only one of these slow-rate state variables may still result in inaccurate or erroneous inferences.

3.1.3 Perceptual access and rational reasoning

General world knowledge and beliefs are acquired by rational agents through the same procedures. More specifically, real-life data from the environment is perceived via perceptual access and processed via rational reasoning (see the rectangular elements in Figure 2) by the agent, and accordingly the agent makes a judgement (rationally perceived knowledge oval in Figure 2). The rationally perceived knowledge may be transformed into a belief or general world knowledge by the rational agent. In the model proposed by Baker (2012) these procedures are represented as a single element called the principle of rational belief (see Figure 1). This simplification was shown to be sufficient to explain the relationship between the environmental inputs and the inferred beliefs in the simple environments that were considered by Baker (2012) in their case studies. In real-life scenarios, however, a more complicated procedure (as explained earlier in this section) occurs before a belief or a piece of general world knowledge is developed based on the raw data from the external environment. In particular, this procedure may be highly personalised for different rational agents. Moreover, real-life data should be distinguished from the data that is perceived by a rational agent. A rational agent may access only a portion of the real-life data in each interaction with its environment. Nevertheless, real-life data may be altered at any given time, regardless of whether the rational agent perceives the data or not. Consequently, rational agents may hold false or inaccurate beliefs (c.f. the Sally-Anne experiment (Baron-Cohen et al. 1985)). Prior research has proven that it is essential to ToM that observer agents recognise that observed agents may hold false or inaccurate beliefs (Wellman et al. 2001; Rabinowitz et al. 2018). To address the discussed aspects, the process that transforms real-life data into beliefs and pieces of general world knowledge is decomposed into smaller, well-defined sub-processes in the proposed cognitive model: the perceptual access and the rational reasoning (see Figure 2). The process of perceiving the real-life data, called *perceptual access*, receives (part of the) real-life data as input and returns the *perceived data* as output.

In practice, both internal and external factors can influence perception of the real-life data. On the one hand, perception depends on rational agents, i.e., when located in the same environment different rational agents may notice or decide to receive different types of data from the environment. It is essential to include this effect in the cognitive model in order to provide more precise interactions among rational agents. The following example illustrates the importance of personalisation of the perception procedure.

Example 4 Suppose that a tourist (i.e., the observed agent) tells the tour leader (i.e., the observer agent) that she has already been to the historical city centre. The tour leader may suppose the tourist has a perfect knowledge of the real-life data, including the church, old building of the City Hall, and all the souvenir shops (general world knowledge of the observed agent according to the observer agent), while in her previous visit the tourist had overlooked the old building of the City Hall because souvenir shops appealed much more to her than old buildings. Then the general world knowledge considered by the tour leader for the tourist is inaccurate. In case the tourist tells her close friend (i.e., an observer agent who has access to personalised perception of the observed agent), instead of the tour leader, that she had once been to the historical city centre, the friend might suppose - based on a personalised perception procedure of the tourist - that she had overlooked the old building of the City Hall.

The above example shows that perceptual access of a rational agent is highly personalised.

On the other hand, this perceptual access is influenced by environmental factors. For example, a rational agent may not receive some environmental visual data because the space is occluded, or might inadvertently hear a sound. Therefore, the element corresponding to the perceptual access is placed in the border of the *agent* box and its environment (see Figure 2). Once real-life data is perceived by a rational agent, the resulting information goes through the *rational reasoning* process, yielding a rationally perceived knowledge, which is then transformed into a belief or a general world knowledge (c.f. Figure 2). The following example demonstrates the importance of decomposing the process that yields the rationally perceived knowledge, as it is suggested in our cognitive model.

Example 5 Brian, Charlie, and Diana are inside a shopping mall, when someone soaked in water enters. Although they cannot directly see the outside, before they entered the shopping mall it was sunny. Brian does not notice the person who is soaked in water (no updated *perceptual access*) and thus, he *believes* that outside is sunny. Charlie and Diana notice the person soaked in water (updated *perceptual access*). Charlie reasons and accordingly *believes* that it must be raining now, while Diana reasons that this person has fallen into a ditch (*different, personalised rational reasoning*) and *believes* that it is still sunny outside.

When an observer agent infers about the beliefs of Brian, Charlie, and Diana (all as observed agents), if the observer agent does not consider personalised rational reasoning for these observed agents, it may infer that the three of them believe that it is now raining outside.

3.1.4 Emotions and personality traits

In both human-human and human-robot interactions, emotion recognition is of utmost importance. More specifically, identifying the emotions of the observed agent enables the observer agent to infer about the overall state-of-mind of the observed agent (Kwon et al. 2008; Saxe and Houlihan 2017), and to select the most appropriate behaviour and interactions accordingly (Tapus and Mataric 2008; Zaki and Craig Williams 2013; Lee et al. 2019). Moreover, the personality traits of a rational agent act as a regulator of their emotions (Bono and Vey 2007). Therefore, incorporating the emotions and personality traits in a cognitive model yields more genuine and engaging human-like interactions for a machine that uses this model (Tapus and Mataric 2008; Leite et al. 2013).

Personality traits are included as slow-dynamics state variables within the proposed cognitive model (see Figure 3). Beside emotions (which are discussed in detail below), goals are directly affected by personality traits. For instance, while an introvert rational agent develops the goal of hiding from strangers, an extrovert rational agent may develop the goal of initiating a talk with a stranger. Although beliefs may also be affected by personality traits, this influence is indirect (this will be covered in more detail in Section 3.1.5).

Next, we discuss the mutual influences of emotions and other state variables in the cognitive model. In particular, we discuss which (combination of) state variables and inputs generate or influence the emotions of a rational agent and how these emotions should be incorporated into the proposed cognitive model.

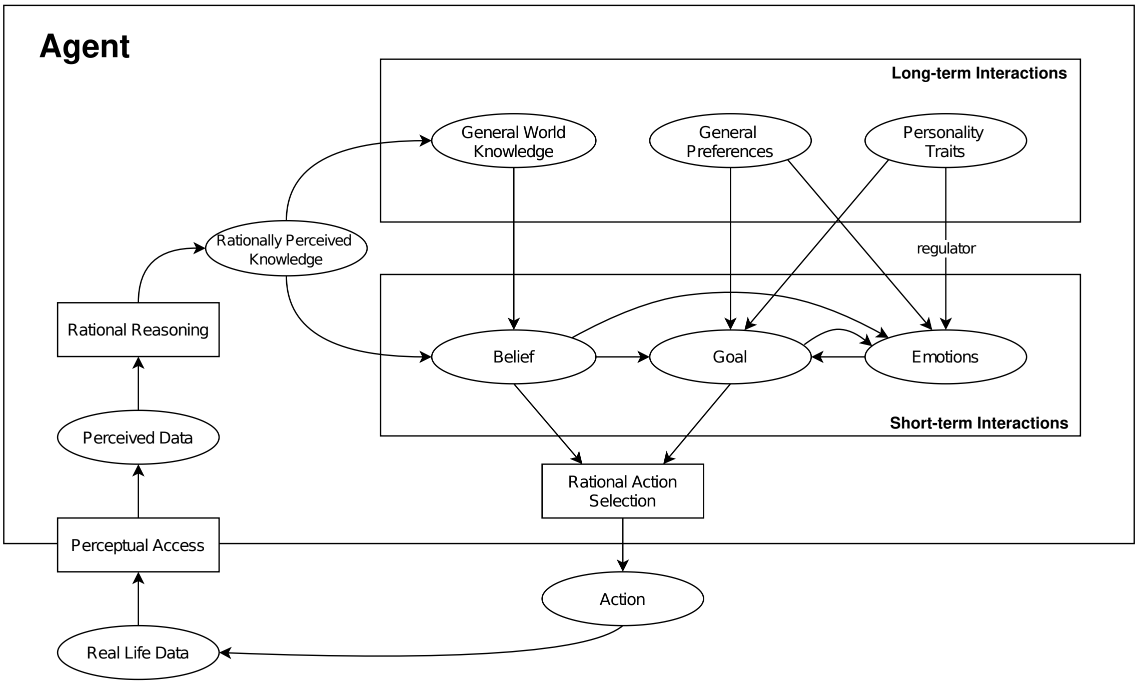


Fig. 3: Proposed cognitive model including the emotions and personality traits.

State variables that influence emotions: Emotions of a rational agent may be stimulated by its beliefs that are generated by the rationally perceived knowledge (see Figure 3). Note that emotions are not directly generated by external inputs (i.e., real-life data received by a rational agent), but by how the rational agent internalises and interprets these inputs. To clarify this, we subsequently give an example.

Example 6 While walking on the street, Elisa's wallet falls out of her purse (*real-life data*). Later on in a shop, Elisa reaches for her wallet and realises that it is not in her purse (*perceptual access*). She reasons that she has lost the wallet (*rationally perceived knowledge*). She then supposes that she has lost her wallet (*inference of a belief based on the rationally perceived knowledge*). This belief makes her anxious (*stimulation of emotions*).

In the given example, before Elisa notices that her wallet is missing (i.e., without *perceptual access*) and reasons that she has lost it (i.e., without *rational reasoning*), she was not anxious (no stimulation of *emotions*). In a different situation, for the same perceptual access that causes the same perceived data, i.e., a missing wallet, Elisa may reason and believe that she has left her wallet on the dining table at home (*different rational reasoning and hence different rationally perceived knowledge*). Therefore, Elisa will not be anxious (no stimulated *emotions*). In summary, independently of what the real-life data is (e.g., the wallet has fallen on the street or is at home) the emotions of a rational agent may be moderated by the perceptual access of the agent to that data and by the reasoning that it applies to the perceived data. In other words, the emotions of a rational agent depend on its beliefs rather than on real-life data directly.

Both goals and general preferences, alongside a belief, can impact emotions. On the one hand, when a rational agent follows a goal and develops a belief that is in line with the fulfilment of that goal, positive emotions may be stimulated. On the other hand, when a rational agent follows a goal and develops a belief that hinders the chances of fulfilling that goal, negative emotions may be stimulated. Similarly, when a general preference is supported by a developed belief, positive emotions can be generated. On the contrary, when developed beliefs conflict with the general preferences of a rational agent, negative emotions may be developed. General preferences mostly influence the emotions of a rational agent indirectly and via generating a goal - that alongside a belief - stimulates emotions. Although excluding the direct influence of general preferences over emotions can simplify the resulting cognitive model, developing a direct connection between general preferences and emotions in some cases is essential. The next two examples show, respectively, the effect of goals and general preferences on the emotions.

Example 7 Frank is exploring a new city for the first time and wants to buy an ice cream (*goal*). While walking, he notices a few people across the street who are eating ice cream (*perceived data*). Correspondingly, he reasons and believes that there should be an ice cream shop close by (*rationally perceived knowledge* transformed into a *belief*), which makes him feel satisfied (*stimulation of emotions*).

This example shows a case where a belief by itself does not stimulate emotions, but the belief together with a goal does. In other words, if Frank did not want to eat an ice cream, the belief that an ice cream shop is nearby would not influence his emotional status. The next example shows a case where general preferences alongside beliefs directly stimulate emotions.

Example 8 Grace is afraid of dogs (*general preference*). While walking in a park, she notices the footprints of a dog (*perceptual access*) and correspondingly reasons and believes that there should be a dog nearby (*rationally perceived knowledge* transformed into a *belief*). This belief makes her anxious (*stimulation of emotions*).

Any direct effect from general world knowledge on the emotions is negligible, since in practice general world knowledge is transformed into beliefs, which influence the emotions directly.

Personality traits of rational agents may influence how much a certain belief affects their emotional state. For instance, compared to an introvert, an extrovert rational agent may experience more excitement for being invited to a social event. Personality traits do not generate emotions by themselves, which is in line with the fact that emotions are fast-dynamics state variables that are temporary and event-triggered, whereas personality traits constantly exist. In other words, if personality traits were directly generating emotions, a rational agent had to continuously experience those emotions. Personality traits instead boost or hinder the emotions and may be seen as regulators of the emotions.

To summarise, three emotion triggers were identified. As explained, a belief is always present when an emotion is triggered. Thus, one of the identified triggers is caused individually by beliefs, like in *Elisa's* example. Secondly, certain combinations of goals and beliefs can lead to an emotional response, as illustrated by *Frank's* example. The third trigger corresponds to the trigger generated by combinations of certain general preferences and beliefs, such as *Grace's* example. This division is important to establish that some variables cannot trigger an emotion *per se* and only certain combinations of variables are able to do so. All these triggers are boosted or hindered by personality traits.

Remark 3 We refer to the emotion trigger caused by combinations of beliefs and general preferences as *emotion trigger 1*. The emotion trigger generated solely by beliefs is named *emotion trigger 2*. Finally, the emotion trigger caused by combinations of beliefs and goals is called *emotion trigger 3*.

State variables that are influenced by emotions: Studies show that emotions can affect the immediate goals and desires of rational agents (Raghunathan and Pham 1999; Andrade and Ariely 2009; Lerner et al. 2015; George and Dane 2016). More specifically, emotions may result in the development of a goal that contradicts the general preferences of a rational agent or in the change of a goal that was previously made by the rational agent. For instance, gratitude can galvanise rational agents into helping others (Lerner et al. 2015), or anxiety may trigger rational agents to avoid stressful situations (Raghunathan and Pham 1999). The influence of emotions over goals is introduced into the proposed cognitive model shown in Figure 3 via a directed link. The following example illustrates the influence of emotions on the developed goals of a rational agent.

Example 9 Hailey has planned to go to a party tonight (original *goal*). In the afternoon, she receives bad news that make her sad (stimulated *emotion*). As a consequence, she decides not to go to the party anymore (*change in the goal due to the emotions*).

This example shows how triggered emotions can affect an already developed goal of a rational agent. Similarly, if Hailey's general preference is to participate in social events, but just before she hears about the party she gets upset by some bad news, she may develop a goal (i.e., skipping the party) that contradicts her general preferences.

While emotions do not directly influence beliefs, they can affect the processes that result in judgements made by a rational agent (Raghunathan and Pham 1999; Andrade and Ariely 2009). More specifically, positive emotions may introduce optimistic biases into the process of generation of new beliefs, whereas negative emotions may lead to the formation of overly pessimistic beliefs (Lerner et al. 2015). It hence makes sense to represent the influence of emotions on the development of beliefs in the proposed cognitive model. This influence will be comprehensively be discussed in Section 3.1.5.

3.1.5 Bias in the belief generation of rational agents

In Section 3.1.4 we mentioned that emotions may affect the formation process of the beliefs of a rational agent. In this section, we elaborate on this idea and on incorporating it into the proposed cognitive model.

Bias in beliefs generated by emotions: Although beliefs may be biased by emotions - which themselves are stimulated by beliefs alone or by other state variables - These beliefs may be independent of those beliefs that contributed to the generation of these emotions. The next examples clarify this further.

Example 10 Igor and Jane are having a walk together. While walking, they see a dog (*real-life data*, which after perceptual access and rational reasoning, results in the belief that there is a dog nearby). Since Igor is afraid of dogs (*general preference* of Igor), he feels afraid (the combination of the developed belief with the general preference stimulates this *emotion*), and starts to believe that the dog might harm him (*belief* of Igor is biased by his emotion). Jane, however, does not feel any fear and hence, does not believe in any threat from the dog (unlike Igor, Jane's *belief* is not biased since there are no emotions involved, although they both had the same initial belief that a dog is nearby).

Bias in beliefs generated by goals: Similarly to emotions, goals of rational agents may introduce some bias into their rational reasoning processes. Next, we give an example of this effect.

Example 11 Kevin is a football fan (*general preference*). The team he supports is currently in the second place in the championship. When all evidences are studied by an objective analyst, they conclude that - although not impossible yet - the chances that Kevin's favourite team wins the championship are very small (*unbiased belief*). Since Kevin wants his team to win (Kevin's *goal*[2]), he believes that his team will win (*belief biased by a goal*).

Similarly to the emotions, the intensity of the bias of the beliefs of a rational agent may depend on the personality traits (i.e., personality traits may regulate - rather than generate - the bias of the beliefs). In order to include the bias in the beliefs[3] of a rational agent in the proposed cognitive model, two options are proposed: (I) The bias is introduced after the rational reasoning process is executed (see Figure 4). This results in the *perceived knowledge* (as opposed to *rationally* perceived knowledge in Figure 3), which develops into the biased beliefs. (II) A more realistic case corresponds to a cognitive model that does not divide the process of developing a (biased) belief into two stages. Instead, a biased belief is directly generated, i.e., a rational agent does not necessarily recognise there is a bias in its reasoning processes. On the contrary, the agent usually considers its beliefs to correspond to reality.

Although case (II) may correspond more to the reality of a rational agent's reasoning, an observer agent is more likely to consider a cognitive model corresponding to case (I) when analysing an observed agent. More specifically, when rational agents reason about the cognition of other rational agents, they consider both the rational beliefs and the biased beliefs (Wang and Jeon 2020), which corresponds to the cognitive model represented in Figure 4. Thus, although the cognitive model in Figure 5 portrays the human's reasoning more realistically, the one illustrated in Figure 4 is more in line with the aims of this paper.

[2] In this paper, the concepts of goals, desires, wishes, and needs of a rational agent are used inter-changeably.

[3] We assume that the general world knowledge of a rational agent is not affected by the *bias* variable. The reason for this assumption is that general world knowledge is a slow-dynamics state variable, while the biases are caused by current emotions or goals, which are fast-dynamics state variables. Assuming that the bias could influence the general world knowledge would go against Remark 1, which states that slow-dynamics state variables are considered fixed parameters and we are not interested in modelling their evolution. This assumption does not imply that the general world knowledge of an agent is always rational, but that the fast-dynamics state variable *bias* does not influence it.

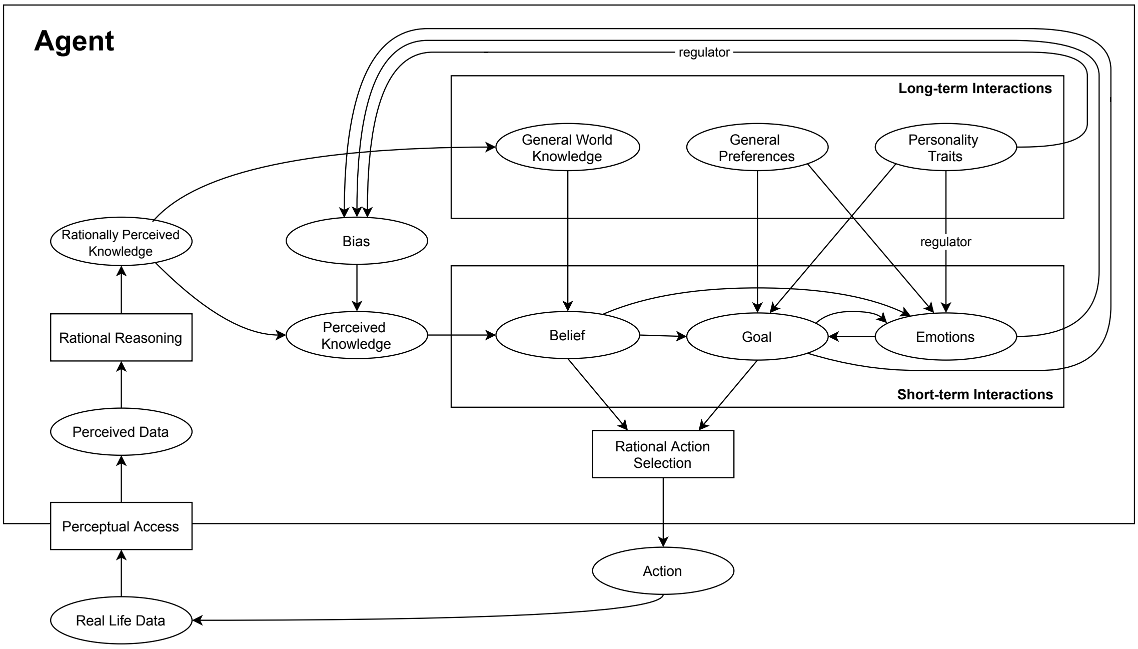


Fig. 4: Cognitive model that incorporates the effect of the bias on the beliefs, considering the bias as an element that alters the rationally perceived knowledge and yields the perceived knowledge. The bias is caused by emotions and/or goals and is regulated by personality traits. Moreover, the bias affects the beliefs, but not the general world knowledge. In case there is no bias, the perceived knowledge is equivalent to the rationally perceived knowledge (this figure will reduce to Figure 3).

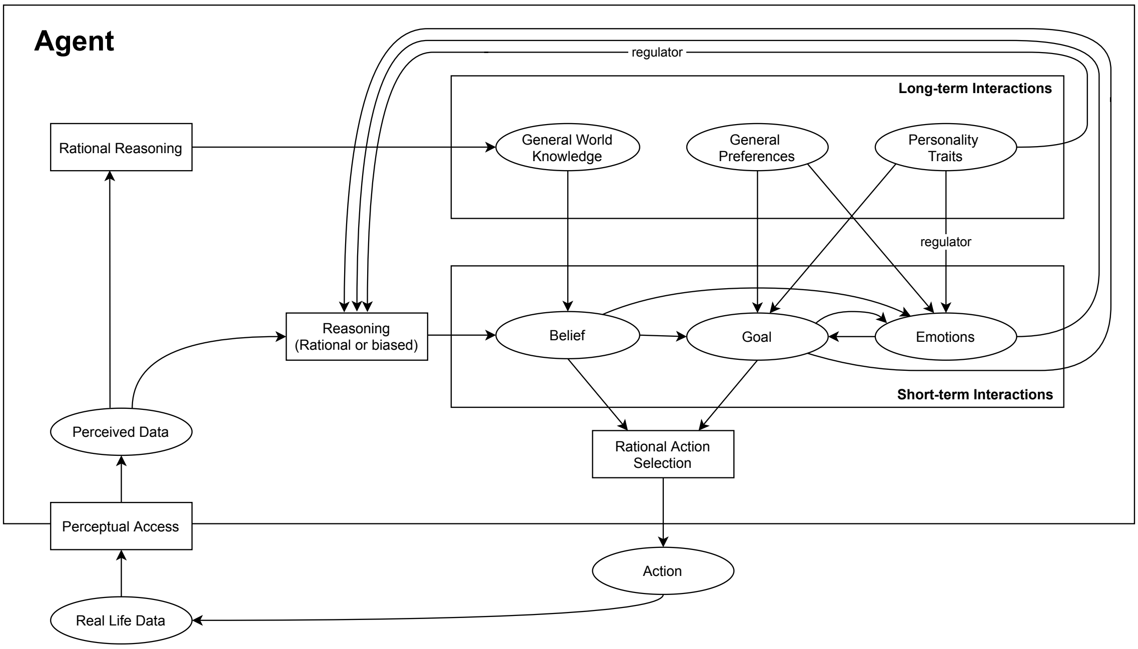


Fig. 5: Cognitive model including the reasoning processes executed by a rational agent on the perceived data: The main difference with Figures 3 and 4 is in considering two process blocks for the rational agent's reasoning; one that executes rational reasoning and transforms the perceived data directly into general world knowledge, and the other one that executes either rational or biased (by emotions and/or goals, and regulated by personality traits) reasoning to develop the beliefs (in case the second block executes rational reasoning, this figure reduces to Figure 3).

3.2 Inverse Inference of Emotions from Actions

According to the principle of rational action, which was described in Section 1.2, actions of rational agents are a direct consequence of their beliefs and goals (Dennett 1987; Baker et al. 2011; Jara-Ettinger et al. 2016; Saxe and Houlihan 2017). Contrarily to the other two fast-dynamics state variables (i.e., beliefs and goals), emotions do not directly generate actions. Nevertheless, as it was explained in Sections 3.1.4 and 3.1.5, emotions contribute directly to the goals and indirectly to the beliefs of rational agents (see Figures 4 and 5) and thus, indirectly to the actions. While Baker et al. (2007, 2011) have discussed the inverse inference of the beliefs and goals from actions, since our proposed cognitive model incorporates a third group of fast-dynamics state variables, i.e., emotions, the inference of the emotions of rational agents from their observed actions is discussed next.

In particular, two general cases are considered where an inverse inference of emotions from actions is possible (and necessary for precise consecutive estimations and/or predictions by the cognitive model). These two cases are based on the following assumptions:

1. Assuming that no emotions are involved, the present model accurately infers the beliefs and goals.
2. Expecting a particular action (which is predicted based on a previously inferred goal-belief pair) and observing a different action imply that either the goal was wrongly inferred or the belief of the rational agent was meanwhile updated.

Goal inferred wrongly: Assuming that the belief has correctly been inferred while the observed action of the rational agent is different from that predicted by the cognitive model, then the goal has been wrongly inferred. Note that among all the state variables that influence the goals, i.e., general preferences, personality traits, beliefs, and emotions (see Figure 3), only the fast-dynamics ones are prone to change in the temporal scales considered in this research. Assuming that the inferred beliefs were in line with the beliefs that resulted in the realised action, emotions are considered as the main reason that the goal of the rational agent was wrongly inferred.

Example 12 I see a friend and smile at them, expecting them to smile back (*expected action*). But they do not smile back and turn away (*observed action*). Thus, they might be upset or mad at me (*inferred emotion*).

Belief updated: Secondly, an unexpected action may have been caused by a belief update which prevented the initially inferred goal to be achieved, additionally triggering a negative emotion and/or surprise. For this circumstance to take place, the initially inferred belief-goal pair must consist in a goal that depends on the belief, i.e., for the goal to be fulfilled, the belief must correspond to the reality.

Example 13 I see Lewis going to the swimming pool with his swimming equipment (*initially observed action*), inferring his goal to swim and his belief that the pool is open. Based on the inferred belief and goal, I can predict that Lewis will stay in the pool for some time (*predicted action*). However, some minutes later, I see him walking in the opposite direction (*observed action*). I can infer that some unexpected event occurred which prevented him to fulfil his goal, such as the pool being closed. Thus, he probably feels disappointed or frustrated (*inferred emotion*).

In the previous example, the agent was convinced that a certain belief was true and, consequently, had a goal which depended on that belief. Nevertheless, he was not able to fulfil that goal, probably due to a belief change. That belief change is likely to have triggered a negative emotion, since it prevented the goal to be achieved. Moreover, the presented examples display how the same situation can trigger different emotions in different agents (Saxe and Houlihan 2017), reinforcing the need for personalisation.

Figure 6 represents the two aforementioned generic scenarios in which emotions can be inferred from actions. Both cases require two different interactions or two different actions within the same interaction to enable the described conclusions: from the first, an initial belief-goal pair or a slow-dynamics state variable is inferred, which leads to the prediction of a subsequent action; in the

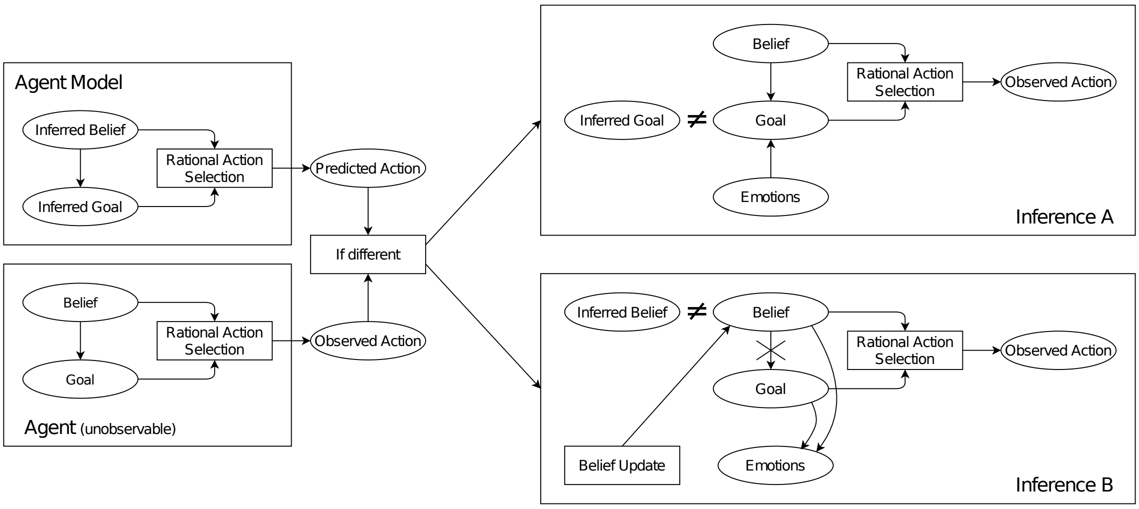


Fig. 6: Emotion inference from actions. When an observed action differs from the action predicted based on a previous inference (left), two scenarios are possible: either the goal was wrongly inferred, which was likely due to the influence of an emotion (top right), or the belief was updated, preventing the agent to fulfil their goal, which causes an emotion (bottom right).

second, the agent executes an action, and the predicted and observed actions are compared. The potential of combining forward and inverse inferences in order to accurately infer a wide range of emotions had already been remarked by Saxe and Houlihan (2017). However, no specific structure was proposed or implemented in this study.

The conclusion that two actions or interactions are fundamental to infer emotions from actions suggests that not only the present goals and beliefs are relevant: it is equally important to keep a record of the past beliefs and goals.

Although emotion generation based on beliefs, goals, general preferences (and moderated by personality traits) was included in the model described in Section 3.1.4, the inferences described in this section and presented in Figure 6 can contribute to a better understanding of the emotions of the rational agent. Furthermore, they enable the correction of goals wrongly estimated or beliefs that have been updated by the agent.

Nonetheless, the foregoing model and the inferences described in this section represent two very different ways to estimate emotions: while the first applies exclusively direct inferences considering the variables which influence emotions, the latter consist in comparing the conclusions taken from two different interactions or actions, combining both inverse and forward inferences.

Thus, the two methods should be implemented in parallel modules, integrating the information given by the forward inferences made by the model formalised in Section 3.1 with the information yielded by the inferences from actions described in this section. More specifically, the latter, which is more specialised, can be used to correct the emotions, beliefs and goals inferred by the first model. By merging the information yielded by the two approaches, a more accurate global estimation of human reasoning and behaviour can be achieved.

4 Cognitive Models of Observed Agents: Formulation

After formalising the cognitive models, the next step is to formulate them according to appropriate frameworks. To facilitate the formulation, the models were broken into five modules (see Figure 7): the *model core*, which contains solely the internal state variables that influence beliefs, goals and emotions; the perceptual access and rational reasoning, as well as the perceived data, are included in the module termed *input processes*; the rational action selection is represented in a block called *output processes*; the influence of the actions of a rational agent on the real world data is formulated in a *world model*; finally, the model that computes inverse inference of emotions from actions, as described in Section 3.2, is represented in a separate, parallel module.

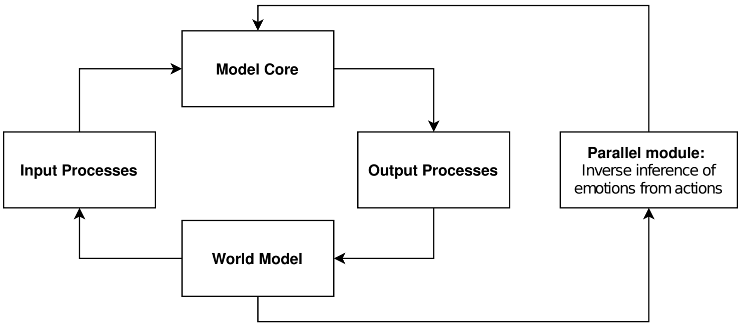


Fig. 7: Division of the models into different modules for formulation. The *model core* only includes the *internal state variables* that influence the beliefs, goals and emotions. The *input processes* include the perceptual access and rational reasoning, as well as the perceived data. The *output processes* module consists only in the rational action selection block. The influence of the actions of an agent on the real world data is formulated in the world model. Finally, the module described in Section 3.2, which handles inverse inference of emotions from actions, is formulated as a parallel module to the model core.

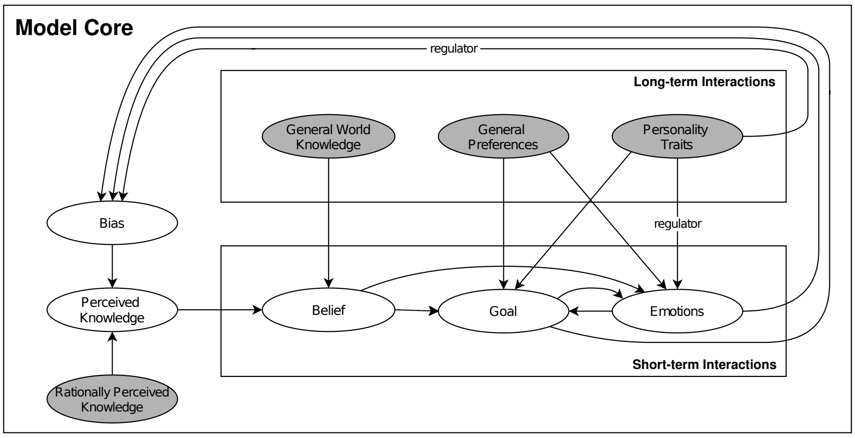


Fig. 8: *Model core* module. The *model core* only includes the *internal state variables* that influence the beliefs, goals and emotions. The elements that are state variables from the perspective of the model core are represented in white, while the elements that are inputs of the model core are represented in grey.

The present formulation focuses mainly on the model core (see Figure 8), since this module is the main contribution of the present project and can be operated on its own. The input and output processes are also considered in the formulation. The additional module of inverse inference of emotions from actions discussed in Section 3.2 is not included in the present formulation, since it can be formulated independently. As mentioned in Remark 1, the evolution of slow-dynamics variables is out of the scope of the present research project and, consequently, is not considered in the present formulation. From the point of view of the model core, the slow-dynamics state variables are inputs.

Taking into consideration the differences between the model core and the input and output processes, these parts of the model can be formulated by different frameworks, as long as the frameworks are compatible with each other.

4.1 Bayesian Networks

Several prior studies (Baker et al. 2011, 2017; Jara-Ettinger et al. 2016; Saxe and Houlihan 2017; Lee et al. 2019) are based on the assumption that the reasoning and behaviours/actions of humans follow a Bayesian framework. More specifically, these studies assume that humans intuitively process the world in terms of probabilities by making connections between different premises and by inferring the likelihood of pieces of knowledge based on their prior knowledge (Johnson-Laird 1994). Moreover, according to Jara-Ettinger et al. (2016) the actions of rational agents can be predicted based on their current state-of-mind (forward inference) and their behaviour and actions serve as an indicator of their state-of-mind (inverse inference).

Bayesian Networks (BNs) are graphical models that describe the causal relationships between a set of n variables $X_1, \ldots, X_n$ in (conditional) probabilistic terms (Heckerman 2008). Given the causal relationship between the elements of the model proposed in Section 3 and based on the state-of-the-art research that considers these relationships to be probabilistic, BNs are promising frameworks for formulating the cognitive model.

Despite the clear correlation between the state-of-mind and the behaviours or actions of humans with the structure of BNs, only few prior research projects have applied this type of networks to represent the aforesaid processes (Lee et al. 2019). This is likely due to the requirement of explicitly defining the conditional probabilities, which can be very challenging in complex and unobservable environments where an inferring agent analyses the state-of-mind or predicts the actions of an observed agent. Furthermore, the structure of a BNs correspond to a direct acyclic graph, meaning that loops should be avoided in the graphical representation of a BN (Heckerman 2008). However, the proposed cognitive model corresponds to a cyclic graph. Thus, in order to implement it using BN framework making the model acyclic - by, for instance, excluding some connections - is inevitable. Given these shortcomings, other frameworks were contemplated, and their benefits and disadvantages compared with the BNs' within the scope of representing the model attained in Section 3.

4.2 Partially Observable Markov Decision Process

Baker et al. (2007) demonstrates that Markov Decision Processes (MDPs) are suitable to represent how rational agents behave according to their goals and external constraints, assuming they select rational actions. Nonetheless, MDPs are based on the assumption that rational agents have complete access to their external environment. However, this is a restrictive assumption that may yield an inaccurate description of human's cognition, since it does not represent that agents can hold false beliefs, as explained in Section 3.1.3. Therefore, prior research works have used Partially Observable Markov Decision Processes (POMDPs) to model the behaviour of rational agents according to a ToM framework (Baker et al. 2011; Rabinowitz et al. 2018; Lee et al. 2019). POMDPs encompass the main principles of MDPs, while incorporating the premise that rational agents may observe only a part of the environment at a time. Consequently, rational agents compute and keep the probability of every state of the state space. Therefore, POMDP is a highly promising framework for dealing with uncertain probabilistic/stochastic environments (Kaelbling et al. 1998; Foka and Trahanias 2007), which is also the case when modelling human's cognition. More specifically, the assumption of, e.g., Johnson-Laird (1994), is that when humans make a cognition about their surrounding world, they think, analyse, and reason in probabilistic terms, i.e. humans intuitively consider the likelihood of an event occurring over other possible outcomes given the observed premises.

Remark 4 In MDPs (including POMDPs) rational agents are assumed to select actions that maximise their expected discounted return (i.e., a weighted sum of the present and the future rewards). In formulation of cognitive models via POMDPs, assigning the discounted return to the action selected by the rational agent is included within the *rational action selection* block.

Every rational agent may have different priorities and interpretations about rewarding situations and the relative importance of future and immediate rewards. Therefore, we personalise the reward function (i.e., a mathematical function that determines the value of the immediate reward that is assigned to the action taken by the agent) and the discount factor (i.e., a parameter that defines the importance of future rewards with respect to the most immediate rewards in the discounted return) in our proposed model formulation.

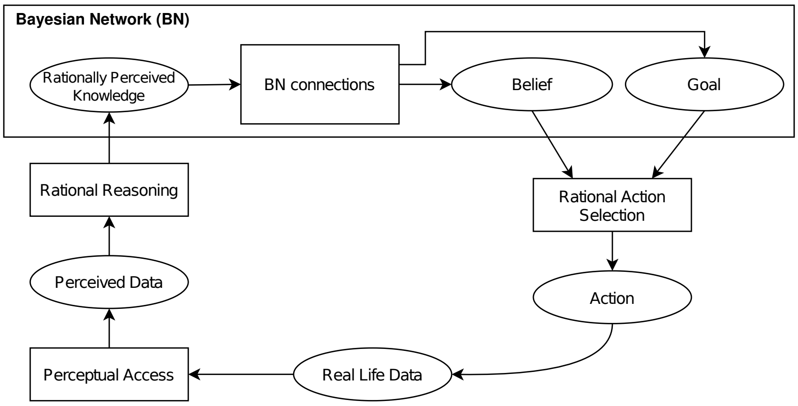


Fig. 9: Elements of the models that are part of the POMDP in our formulation of the model core with a BN, and their connection with the BN.

Finally, in a POMDP framework, rational agents need to keep a belief base that, for all feasible states, encodes the probability of having a particular state at a particular time. The belief base is updated after every interaction of the agent with the environment based on the new observations of the agent (Foka and Trahanias 2007). Note that in the proposed POMDP formulation for our cognitive models, the belief base is formulated as the rationally perceived knowledge. The function of the POMDP that estimates the belief base considering the observations, is formulated as the rational reasoning. The states of the POMDP are formulated as the real-life data, and the observations are formulated as the perceived data. The observation function is formulated as the perceptual access block.

POMDP is used to formulate part of the model (see Figure 9). In order to use this framework and exploit its benefits, it is necessary to represent the variables' links that are not yet included in the POMDP (the model core), connecting the rationally perceived knowledge and the rational action selection elements. Considering the benefits of BNs already stated in Section 4.1, in the formulation that uses POMDPs, the model core is represented by this framework.

In order to ensure that the network is acyclic, the arcs that generate the bias are left out. Similarly, one of the two links between emotions and goals must be excluded (see Figure 10). Firstly, the module that infers emotions from actions already considers the influence of emotions over goals, and updates both variables accordingly. Secondly, the effect of goals over emotions is not considered in any parallel module. Thus, the first influence is excluded from the BN, while the second is comprised.

The main difference between heretofore formulations of POMDPs and the one proposed in this research project lies in the reward function indirectly depending on the belief base, rather than directly. This indirect influence is mediated by the BN and encompasses all the other internal state variables. The network's joint probability is given by:

$$\begin{aligned} p(GWK, GP, PT, RPK, Bias, PK, B, G, E) = \\ = p(GWK) \cdot p(GP) \cdot p(PT) \cdot p(RPK) \cdot p(Bias) \cdot p(PK|RPK, Bias) \\ \cdot p(E|B, G, GP, PT) \cdot p(B|PK, GWK) \cdot p(G|B, GP, PT) \end{aligned} \tag{1}$$

where GWK stands for general world knowledge, GP for general preferences, PT for personality traits, RPK for rationally perceived knowledge, PK for perceived knowledge, B for beliefs, G for goals and E for emotions.

The probabilities of the variables without parent nodes are given either by the POMDP ($p(RPK)$), or by other modules ($p(Bias)$). As mentioned in the beginning of Figure 8, the slow-dynamics are considered inputs of the model core. The reward function is then computed based on the beliefs and goals provided by the BN, rather than on the rationally perceived knowledge.

Finally, the conclusions reached in Section 3.1.4 regarding emotions being caused by one of three triggers and regulated by personality traits must be comprised in the network (see Remark 3 for the definitions of the emotion triggers). This information can be encoded in the conditional

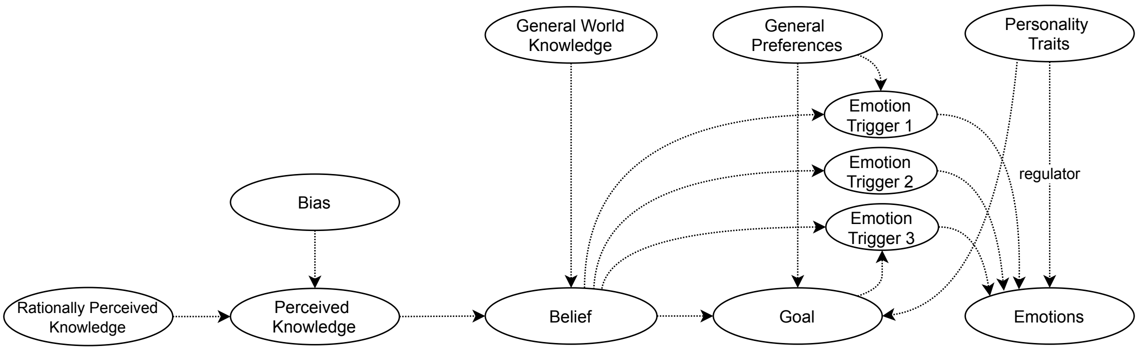


Fig. 10: Formulation of the model core with a Bayesian Network. This network explicitly encompasses the three emotion triggers. For clarity purposes, only the BN is presented, since it is the only part of the model altered by the introduction of the aforementioned trigger nodes. The remaining model and its interaction with the BN are the same as described in Figure 10.

probability of the emotions given its parents $p(E|B, G, GP, PT)$. Alternatively, the three triggers can be explicitly distinguished, as displayed in Figure 10. By representing these nodes, the information that only certain combinations of the emotions' parent nodes can cause emotions is encoded in the network. Consequently, the conditional probabilities of the emotions and *emotion trigger 1 to 3* given their parent nodes are more transparent and easier to define. The network joint probability is given by:

$$\begin{aligned} p(GWK, GP, PT, RPK, Bias, PK, B, G, E, E_{t1}, E_{t2}, E_{t3}) = \\ = p(GWK) \cdot p(GP) \cdot p(PT) \cdot p(RPK) \cdot p(Bias) \cdot p(PK|RPK, Bias) \\ \cdot p(E_{t1}|B, GP) \cdot p(E_{t2}|B) \cdot p(E_{t3}|B, G) \cdot p(E|E_{t1}, E_{t2}, E_{t3}, PT) \\ \cdot p(B|PK, GWK) \cdot p(G|B, GP, PT) \end{aligned} \tag{2}$$

where E_{t1}, E_{t2}, E_{t3} stand for the emotion triggers 1, 2 and 3, respectively.

4.3 Fuzzy Cognitive Maps

Fuzzy Cognitive Maps (FCMs) are an alternative to BNs for representing the model core. FCMs are a framework often employed to model complex or uncertain systems that can be easily described resorting to human knowledge. An FCM is composed by the concepts that are relevant to describe the system, as well as by the interaction between these concepts, i.e., how each one of the concepts influences the others. Contrarily to BNs, FCMs support cyclic connections (Stylios and Groumpos 2004), which is an added benefit given the structure of the proposed model.

The elements that are relevant to explain the operation of the system are termed *concepts*. The i^{th} concept of the system is represented in the FCM by the cognitive concept C_i. In the present formalisation, the concept C_i is represented by the fuzzy variable A_i. We define $\mathbb{C}$ as the set of all concepts of an FCM and $\mathbb{A}$ as the set of all possible realisations of every concept. As mentioned, FCMs comprise the influence between variables. The intended directed influence of a concept C_i over another concept C_j is defined as a *linkage* (see Figure 11). Every linkage is described by a degree of causality and, consequently, has a weight $w_{ij} \in [-1; 1]$ that reflects how much C_i influences C_j. If w_{ij} is positive, then an increase of A_i implies an increase of A_j. Contrarily, if w_{ij} is negative, an increase of the value of A_i leads to a decrease of the value of A_j. If w_{ij} is null, a change in the value of A_i does not influence A_j. Furthermore, the larger the absolute value of this weight, the larger the influence of A_i over A_j.

In the original formulation of the FCM, the weights are constant. Nonetheless, to accurately model most real-world systems with an FCM, it is necessary to consider variable weights (Carvalho and Tomé 2001; Mourhir et al. 2016). In rule based FCM, a framework introduced by Carvalho and Tomé (2001), the value of a weight w_{ij} depends on the value of the causing variable A_i. In the present system, the weights at time step $k + 1$ of some linkages depend on the values of the

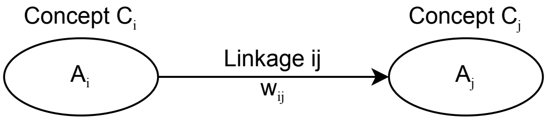
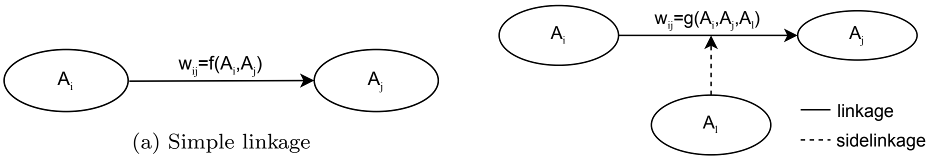


Fig. 11: FCM basic structure: concepts and linkages



(a) Simple linkage

(b) Complex linkage

Fig. 12: Simple and complex linkage

causing concept A_i at time step k, the affected concept A_j at time step k, or even another concept A_ℓ at time step k.

To address the needs of the present system, in this research we propose and define the concepts of *side linkages*, *simple linkages* and *complex linkages*, and introduce a different mathematically representation of the linkages. A *side linkage* ij, ℓ shows the directed influence of a concept C_ℓ on a linkage ij (see Figure 12a). Given that not all linkages are influenced by a side linkage, it is important to distinguish between linkages that are influenced and not influenced by a side linkage. The linkages that are not influenced by a side linkage are called *simple linkages* (see Figure 12a). The group of a linkage that is influenced by one or several side linkages and those side linkages is called a *complex linkage* (see Figure 12b).

A simple linkage between concepts C_i and C_j is given by an ordered pair (i, j). The weight of a simple linkage is then given by function $f : \mathbb{A}^2 \to [-1; 1]$. The set of all the ordered pairs that correspond to simple linkages is called $\mathbb{L}$. Similarly, a complex linkage that is composed by the linkage between concepts C_i and C_j and the side linkage with origin in concept C_ℓ is given by the ordered trio (i, j, ℓ). The weight of the complex linkage is computed by $g : \mathbb{A}^3 \to [-1; 1]$. The set of all ordered trios that correspond to complex linkages is called $\overline{\mathbb{L}}$.

The generalised equation to update the resulting FCM is given by:

$$\begin{aligned} A_j^{k+1} = h\Bigg(& \sum_{\{\forall i | (i,j) \in \mathbb{L}\}} f(A_i^k, A_j^k) \cdot A_i^k \\ & + \sum_{\substack{\{\forall i \; \exists \ell | (i,j,\ell) \in \overline{\mathbb{L}}\} \\ \{\forall \ell | (i,j,\ell) \in \overline{\mathbb{L}}\}}} g(A_i^k, A_j^k, A_\ell^k) \cdot A_i^k \\ & + w_{jj} \cdot A_j^k \Bigg), \text{ for } A_j \in \mathbb{A} - \mathbb{A}_i \end{aligned} \tag{3}$$

where k is the time step, h is a threshold function that bounds the output to the interval defined for A_j, w_{jj} defines how much the value of A_j^k influences the value of A_j^{k+1}, and $\mathbb{A}_i$ is a subset of $\mathbb{A}$ that defines the concepts that are inputs. The choice of the threshold function h must be done taking into account the interval considered by A_j. Moreover, the concepts that are considered inputs cannot be updated: even if they do not have linkages that influence them (which implies that the first two terms of Equation 3 are nonexistent), and $w_{jj} = 1$ their value would not remain constant if they were updated, due to the threshold function (Equation 3). In other words, if Equation 3 is used to update the concepts that are inputs, their update would be given by:

$$A_j^{k+1} = h(A_j^k) \tag{4}$$

which would imply that A_j would not be constant if this concept was updated according to Equation 3, since $h(x) \neq 1$.

When constructing the FCM, for each linkage (simple or complex), there is the need to define the functions f and g. Moreover, several approaches can be used to define such functions, such as using crisp functions, or fuzzy inference systems similarly to what is done in rule based FCMs (Carvalho and Tomé 2001).

When representing the model developed in Section 3 using an FCM, the *concepts* of the FCM correspond to the cognitive states of the model, and the *linkages* and *side linkages* correspond to directed connections of the model. In FCMs, the concepts can be bounded to the interval $[0, 1]$ or $[-1, 1]$. Given that most mental states can be represented by dichotomies, the second interval is more suitable for the present concepts. Consequently, the threshold function mentioned in Equation 3 is selected to be $h(x) = \tanh(x)$, since it bounds the output to $[-1, 1]$.

4.4 Framework Choice

The framework or combination of frameworks chosen to represent the model depended on the framework that was the most suitable to formulate the model core.

As explained, both BNs and FCMs can be used to formulate the model core, since they both have the capability of representing a system composed of several variables, and the connections between those variables. Nonetheless, BNs are acyclic graphs, which would require modifying the current model in order to represent it with a BN. Moreover, implementing the model with a BN would require defining the probability functions for each connection. Contrarily, FCMs offer a much more intuitive way to identify the connections between variables: for each connection (linkage), the effect of varying the causing variable on the affected variable is studied, and a weight is attributed to it. For an intuitive system such as the human cognition, the process of attributing weights to the linkages rather than defining conditional probability functions for each linkage is more simple and accurate.

Although an extensive search was carried out, no comparison between the computational efficiency of POMDPs and FCMs was found. There are studies that compare these two frameworks, which is natural given their similar structure and application (Sedki and Bonneau de Beaufort 2012; Cheah et al. 2008). Nonetheless, these studies perform qualitative comparisons between the two frameworks, and none of them addresses computational efficiency. Therefore, the computational efficiency will not be taken into account for the present choice of framework.

All in all, FCMs offer two advantages over BNs that are crucial for the implementation of the present model: an intuitive way to formulate the connections between variables in mathematical terms, as well as the possibility of keeping the original, cyclic structure of the model. Furthermore, although the literature agrees that fuzzy logic is suitable to represent the cognition of humans in the domain of SARs (Kahraman et al. 2020), the application of this technique on the present domain has remained unexplored. Therefore, this technique was selected.

Finally, since the POMDPs are based on probabilistic functions, other decision-making tools must be used to formulate the rational action selection, perceptual access and rational reasoning processes.

5 Cognitive Models of Observed Agents: Implementation

To implement and test the model using an FCM, a specific real-life situation was selected, and the model was implemented in MATLAB. In a first phase, the results were qualitatively analysed. In a second phase, an online survey was conducted to collect data regarding the cognition of humans in several circumstances of a real world situation. The predictions made by the model regarding the cognition of the survey's participants were compared with the answers given by those participants in order to validate the model.

5.1 Implementation Setup

The model developed in this project was never previously tested, and a ToM model has never been implemented with an FCM. Given that the input and output processes can be independently represented by a decision-making framework, we focused on implementing and testing only the model core. Therefore, for the implementation, we assumed that the rationally perceived knowledge takes the same value as the real-life data. Moreover, the slow-dynamics state variables were initially left out of the model implementation. Once the part of the model that did not include the slow-dynamics state variables and the emotion trigger 1 was analysed, the slow-dynamics state variables were included.

The following procedure describes the implementation of the model for a real-life situation:

1. The real-life situation, which has a similar structure to the real-life examples presented in Section 3 (e.g., Example 9), is defined.
2. The concepts that play a role in the real-life situation (e.g., beliefs, goals, emotions, biases) are identified. Each concept is associated with an index i. Considering the interval $[-1; 1]$ chosen in Section 4.3 for the variables of the FCM, three linguistic terms are associated to the minimum value ($A_{i_{\min}} = -1$), median value ($A_{i_{\text{med}}} = 0$), and maximum value ($A_{i_{\max}} = 1$) of the range of every A_i. For example, for a certain *goal*, the linguistic terms can correspond to "agent does not desire *the goal*" when $A_i = -1$, "agent does not have a preference regarding *the goal*" when $A_i = 0$, and "agent desires *the goal* very much" when $A_i = 1$. The realisations of A_i between the three aforementioned values intuitively correspond to intermediate linguistic terms to the ones presented.
3. Subsequently, it is necessary to define the functions f and g that yield the weights of every linkage comprised in the model. For each linkage represented in the model, the effects caused by increasing and decreasing A_i on A_j are studied. For the cases in which this effect is constant, the linkage is a simple linkage and the effect is classified as one of the following thirteen linguistic terms: {'Direct positive', 'Very strong positive', 'Strong positive', 'Average positive', 'Weak positive', 'Very weak positive', 'Null', 'Very weak negative', 'Weak negative', 'Average negative', 'Strong negative', 'Very strong negative', 'Direct Negative'}. Each one of the linguistic terms corresponds to a default weight, as described in Table 1. For the cases where the influence of A_i over A_j depends only on these two variables, the effect of the simple linkage is defined as a function of A_i and A_j. More specifically, for each interval of A_i and/or A_j where the influence of A_i over A_j is different, this influence is specified with one of the thirteen aforementioned linguistic terms. Thus, the function f can be represented as a piecewise function. For example, $f(A_i^k, A_j^k)$ can output the weight -0.1, which corresponds to a 'very weak negative' influence (see Table 1), when $A_i^k \leq 0$ and output -0.5, which corresponds to an 'average negative' influence, when $A_i^k > 0$. Finally, in the cases in which the effect cannot be described as a function of only A_i and A_j, a complex linkage is present, and the variable A_ℓ (the origin of the side linkage) is identified amongst the other variables that influenced variable A_j. Subsequently, for each interval of A_ℓ (as well as A_i and A_j, if applicable) where the influence of A_ℓ over A_j is different, this influence is specified with one of the thirteen aforementioned linguistic terms. Hence, the function g can be a piecewise function, if required. For example, if the influence of A_ℓ over A_j is 'null' when $A_\ell \leq 0$ and 'weak positive' when $A_\ell > 0$, then $g(A_i^k, A_j^k, A_\ell^k)$ outputs 0 when $A_\ell \leq 0$ and output 0.25 when $A_\ell > 0$.
4. A set of initial conditions that are relevant to test the model in the chosen situation are defined. These initial conditions are defined by the initial values of the inputs of the model core (e.g., rationally perceived knowledge, slow-dynamics state variables) as well as the initial values of some fast-dynamics state variables. The concepts that define the initial conditions depend on the real-life situation. For example, in some real-life situations it is relevant to consider an initial goal, whereas in others it is not. The set of initial values that are relevant to test the model in the real-life situation defines a scenario.
5. Finally, the FCM must be updated for as many time steps as needed for the variables in $\mathbb{A}\backslash\mathbb{A}_i$ to converge, where $\mathbb{A}_i$ is the set of inputs of the model core.

Linguistic Term	Null	Very Weak Pos./Neg.	Weak Pos./Neg.	Average Pos./Neg.	Strong Pos./Neg.	Very Strong Pos./Neg.	Direct Pos./Neg.
Weight	0	±0.1	±0.25	±0.5	±0.75	±0.9	±1.0

Table 1: Relationship between linguistic variables that describe a linkage and the weights associated to them. Depending on whether the linguistic term includes the term positive or negative, the corresponding weight was positive (Pos.) or negative (Neg.), respectively. Only the term null, which corresponds to a non-existent influence, does not have a positive and negative variant.

5.2 Real-life Situations for Implementation

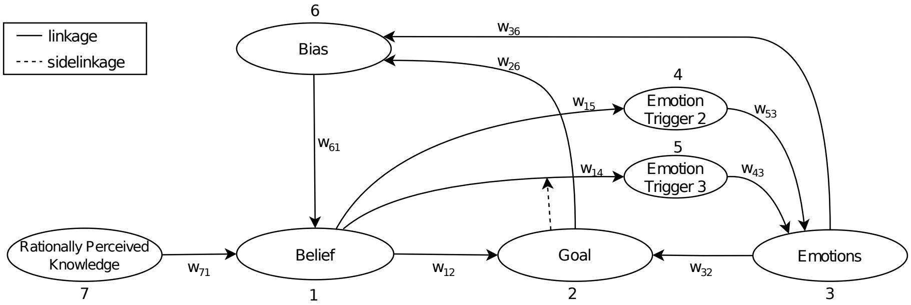
All the examples that were used to formalise the model in Section 3 were considered as starting points to define the ideal real-life situation to implement and assess the model. It is essential that the examples that are chosen to generate the real-life situation comprise all the variables that are present in the model core. In other words, all the variables of the model core, as well as the linkages between them must be relevant to explain at least one of the examples that are used to generate the real-life situation. Therefore, examples 6, 7, 8, and 11 were used as starting points to define the real-life situations used to implement and test the model. The examples of how beliefs, combinations of beliefs and goals, and combinations of beliefs and general preferences create emotions (which correspond to examples 6, 7 and 8) are essential since they comprise all the mentioned variables, as well as the three emotion triggers (see Remark 3 for the definition of the emotion triggers). Moreover, defining a real-life situation based on example 11 is essential to include the reasoning bias. Henceforth, the following real-life situation was defined:

Real-Life Situation 1: An agent has a certain level of desire of achieving the *goal* to do an outdoor activity. The agent looks at the weather forecast (*rationally perceived knowledge*) and forms their own *belief* regarding the weather, which can be influenced by the *bias.* Moreover, the belief of the weather condition, as well as this same belief associated with the goal of the agent, can trigger an emotion through *emotion trigger 2* and *emotion trigger 3*, respectively.

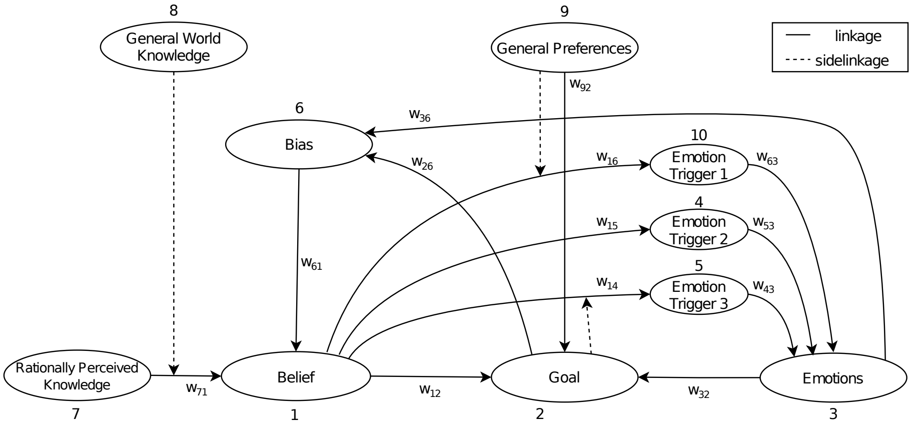
After defining the *real-life situation 1*, the subsequent steps of the implementation procedure were followed. The definition of the concepts, as explained in the second point of the implementation procedure, can be found in Table 2. The structure of the FCM that resulted from the third point of the procedure, i.e., the network of simple and complex linkages between concepts, is graphically represented in Figure 13a. Given that the conversion of *rationally perceived knowledge* into *general world knowledge* is not considered in our formulation and implementation (as explained in Section 4), there is no need to represent both the *rationally perceived knowledge* and the *perceived knowledge* in the FCM. For the sake of simplicity, we implement the bias effect directly on the belief, and do not include the perceived knowledge concept in the FCM. For each simple and complex linkage, the functions f and g were defined. Finally, according to the fourth step in the procedure, the rationally perceived knowledge (input) and the initial goal of the agent (fast-dynamics state variable) were the concepts selected to define the initial conditions. Once the results were analysed (as it will be explained in Section 5.3.1), this real-life situation was extended to include the slow-dynamics state variables, and the following real-life situation was defined:

Real-life Situation 2: Starting from the *real-life situation 1*, the agent additionally has a general preference towards the outdoor activity and knowledge regarding the accuracy of the source of the weather forecast (*general world knowledge*). Moreover, the combination of the *general preferences* and the *beliefs* triggers an emotion through *emotion trigger 1.*

A new FCM that comprised the entire model core was developed based on this situation. Similarly to the *real-life situation 1*, the concepts of the FCM that was used to implement the *real-life situation 2* can be found in Table 2. The *personality traits* were not explicitly included in the FCM as a concept. Instead, they were implicitly included in the choice of the weights. According to the model formalisation, the variable *personality traits* regulates the creation of emotions by other variables, the strength of the goals and the creation of the bias by emotions and goals. Considering that the linkages of an FCM regulate the influence of the concepts on each other, the personality traits are mathematically represented in the FCM by the linkages rather than by a concept.



(a) Graphical representation of the FCM representing *real-life situation 1.*



(b) Graphical representation of the FCM representing *real-life situation 2.*

Fig. 13: Graphical representations of the FCMs that were used to implement the two real-life situations that were defined to test the model. These graphical representations show the linkages and side-linkages between all the concepts. In each iteration, the weights corresponding to every simple and complex linkage w_{ij} are given by the functions f and g, as explained in Section 4.3.

The network of simple and complex linkages between the concepts, as well as the functions that mathematical represent the simple and complex linkages f and g were defined. Figure 13b shows the structure of the resultant FCM. Finally, according to the fourth step of the procedure, the rationally perceived knowledge, the general preferences, the general world knowledge (inputs), and the initial goal of the agent (fast-dynamics state variable) were the concepts selected to define the initial conditions. The FCM was run in different conditions and the results were qualitative analysed.

Concept	**Index**	**Linguistic term**			**Real-life Situation**
		Minimum $A_i = -1$	**Median** $A_i = 0$	**Maximum** $A_i = 1$	
Belief	1	There will be heavy rain	Agent does not know how the weather will be	It will be very sunny	1
Goal	2	Agent does not want to do the outdoor activity	Agent does not have a preference regarding doing the outdoor activity	Agent wants to do the outdoor activity	1
Emotion	3	Very sad	No emotion	Very happy	1
Emotion Trigger 2	4	Very sad	No emotion trigger	Very happy	1
Emotion Trigger 3	5	Very sad	No emotion trigger	Very happy	1
Bias	6	There will be heavy rain	No bias	It will be very sunny	1
Rationally Perceived Knowledge	7	There will be heavy rain	No information	It will be very sunny	1
General World Knowledge	8	Weather prediction is very inaccurate	Weather prediction is mildly accurate	Weather prediction is very accurate	1 and 2
General Preferences	9	Agent strongly dislikes the outdoor activity	Agent does not have a preference regarding the outdoor activity	Agent strongly likes the outdoor activity	1 and 2
Emotion Trigger 1	10	Very sad	No emotion trigger	Very happy	1 and 2

Table 2: Definition of the concepts that comprise the FCMs of the *real-life situations 1 and 2*. Each concept is defined by an index and the linguistic terms that correspond to the bounding values and the median value. The last column defines which concepts are present in the FCM for *real-life situation 1* and which are present in both.

5.3 Model Identification and Assessment

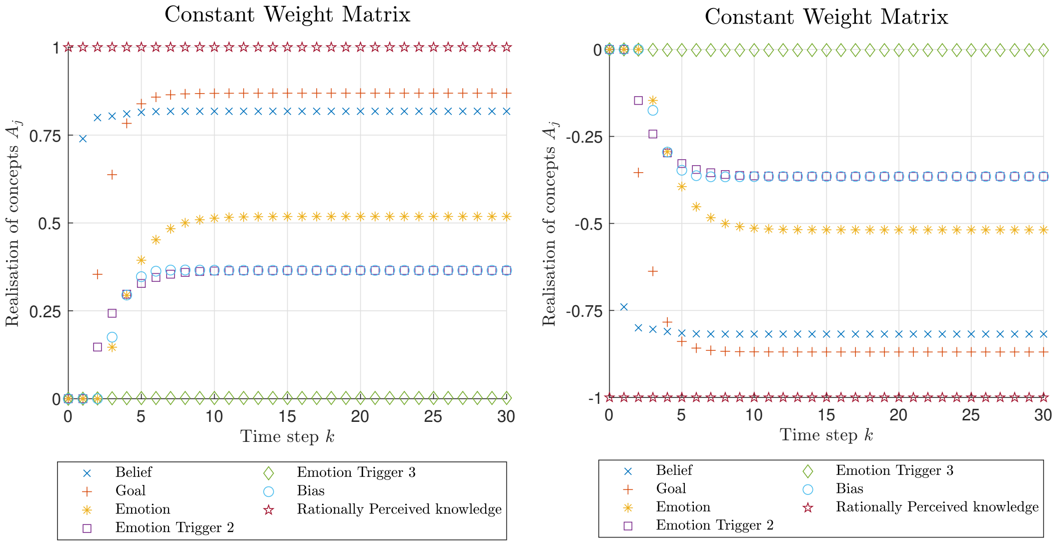
Once the real-life situations were defined and the FCMs were implemented, the time responses of the model to several initial conditions were analysed. Posteriorly, the online survey was carried out and the models that described the cognition of each participant were identified. Finally, the performance of the models was assessed by comparing the answers given by the participants regarding their mental states with the predictions made by the model.

5.3.1 Qualitative assessment of the model

As previously mentioned in Section 5.1, for each real-life situation, there are certain concepts that define the initial conditions. A set of several initial conditions is created in the following way: each concept that defines the initial conditions takes a value in the set $\{-1; 0; 1\}$. Therefore, the number of initial conditions that are present on the set is given by 3^{n_c}, where n_c is the number of concepts that define the initial conditions. For each initial condition on this set, each FCM was updated for 30 time steps, which was enough for the realisation of every variable to converge. Subsequently, the evolution of the realisations of the variables over discrete time was qualitatively studied for the initial conditions on the set. Two types of comparisons were done. The discrete time responses of two different models to the same initial conditions were compared, and the discrete time responses of the same model to different initial conditions were compared. The results that allowed to take the most relevant conclusions are presented and discussed.

As explained in Section 5.2, in the *real-life situation 1*, the initial values of the rationally perceived knowledge and of the goal of the agent defined the initial conditions. Therefore, nine scenarios were tested. These nine scenarios correspond to the combinations of the two concepts taking each one of the three initial values $\{-1; 0; 1\}$. The FCM that corresponds to the first real-life situation was updated for 30 time steps, for each scenario.

The analysis of these scenarios showed that weights should not be constant given that the weights associated to some linkages ij or ij, ℓ depend on the realisation of the concepts A_i, A_j or A_ℓ. Firstly, the weights associated to some linkages ij at time step $k + 1$ depend on the value



(a) Evolution of the FCM variables over 30 time steps, with a null initial goal ($A_2^{k=0} = 0$) and a maximum rationally perceived knowledge ($A_7^{k=0} = 1$).

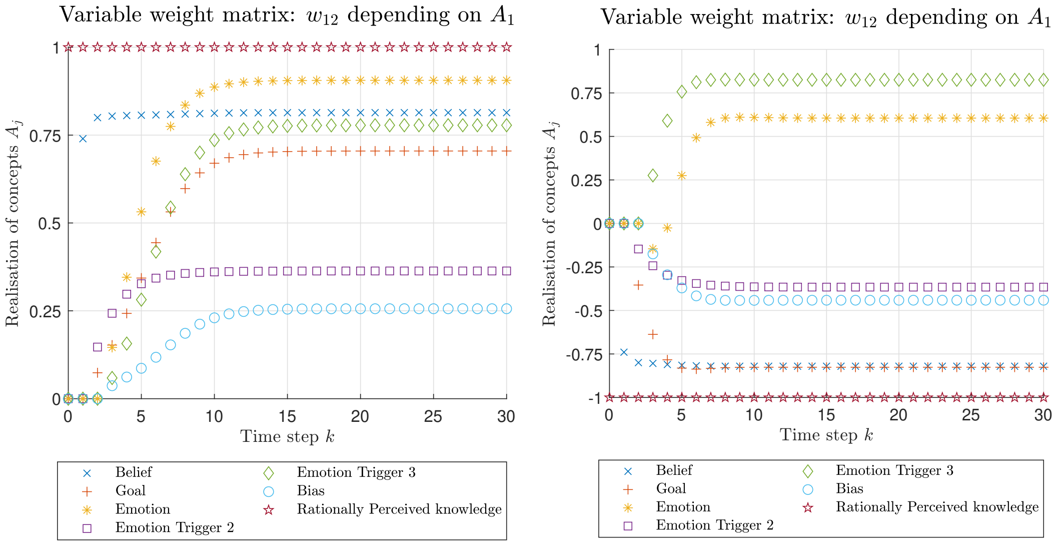
(b) Evolution of the FCM variables over 30 time steps, with a null initial goal ($A_2^{k=0} = 0$) and a minimum rationally perceived knowledge ($A_7^{k=0} = -1$).

Fig. 14: Evolution of the FCM variables over 30 time steps, for two different scenarios, using an FCM in which the weight matrix is constant.

of A_i^k or A_j^k. For example, the weight associated to the linkage between the belief and goal w_{12} depends on A_1, since a negative belief, has a stronger effect on the goal than a positive belief of the same magnitude. Intuitively, it is more likely that the belief that the weather is bad (negative belief) leads the agent to not want to do the outdoor the activity (goal becoming negative), than the belief that the weather is good (positive belief) leads the agent to want to do the outdoor the activity (goal becoming positive). To illustrate this, two scenarios were selected. In both scenarios, the initial goal is null ($A_2^{k=0} = 0$). In linguistic terms, this means that initially the agent does not have a preference about doing the activity. In the first scenario, the rationally perceived knowledge is maximum ($A_7^{k=0} = 1$), implying that the weather is very good, while in the second scenario the rationally perceived knowledge is minimum ($A_7^{k=0} = -1$), meaning that the weather is very bad. The evolution of the concepts in these two scenarios was computed using an FCM in which the weight matrix is constant (see Figure 14) and another in which the influence of beliefs over emotions is stronger for negative values of the beliefs than for positive ones, i.e., the weight associated to the linkage 12 is updated based on A_1 (see Figure 15). Equation 5 represents the function f that defines the linkage 12. The value of w_{12}^- is higher than the value of w_{12}^+, in order to accurately represent the influence of beliefs over goals.

$$w_{12} = f(A_1, A_2) = \begin{cases} w_{12}^- & \text{when } A_1 < 0 \\ 0 & \text{when } A_1 = 0 \\ w_{12}^+ & \text{when } A_1 > 0 \end{cases} \tag{5}$$

The model with the constant weight matrix yields a similar final value of the concept of the goal, in absolute terms, in both scenarios. Therefore, according to this FCM, if the agent does not have a goal regarding doing the outdoor activity and the weather is very good (see Figure 14b), the final goal of the agent is as strong - although positive - as if the agent does not have an initial goal to do the activity and the weather is very bad (see Figure 14a) - although negative. However, the influence of the belief held by the agent about the weather condition on the goal of the agent must be stronger if the weather is very bad (negative belief) than if the weather is very good (positive belief). The model with a constant weight matrix is not able to capture this nuance, while the model that considers the influence of A_i over w_{ij} is able to: the magnitude of the final realisation of the goal is larger when the belief about the weather is negative (see Figure 15b) than when the belief about the weather is positive (see Figure 15a).



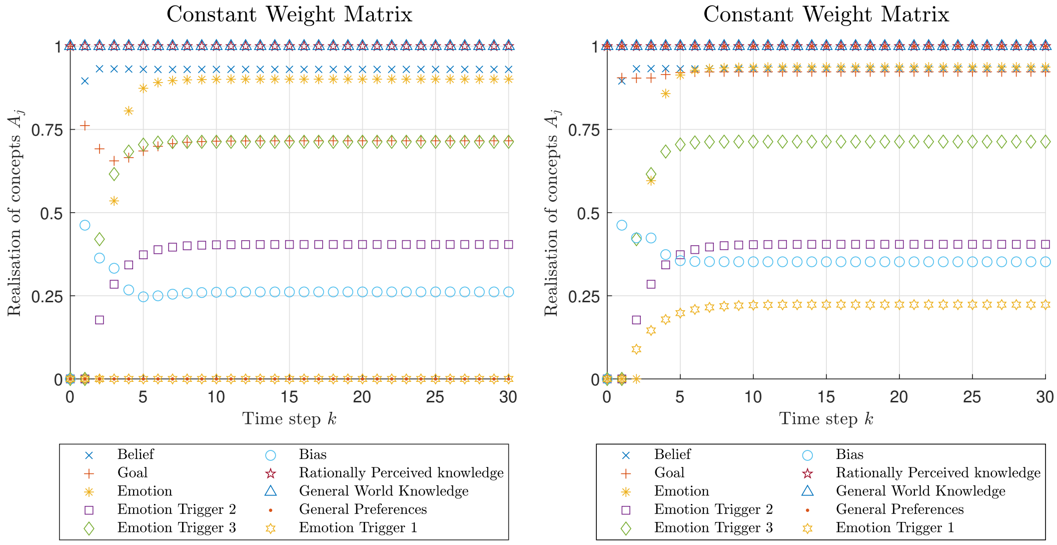
(a) Evolution of the FCM variables over 30 time steps, with a null initial goal ($A_2^{k=0} = 0$) and a maximum rationally perceived knowledge ($A_7^{k=0} = 1$).

(b) Evolution of the FCM variables over 30 time steps, with a null initial goal ($A_2^{k=0} = 9$) and a minimum rationally perceived knowledge ($A_7^{k=0} = -1$).

Fig. 15: Evolution of the FCM variables over 30 time steps, for two different scenarios, using an FCM in which the weight associated to the linkage 12 depends on the value of the belief A_1.

Secondly, the performed qualitative analysis showed that complex linkages are essential to explain the influence of some variables over others, i.e., that the weight associated to some linkages $i\,j, \ell$ depends on the value of another variable A_ℓ. To illustrate this conclusion, two FCMs were considered: the first FCM comprised a constant weight matrix, while the second assumed that the weight associated to the linkage $1\,4, 2$, which is represented as a complex linkage in Figure 13a and Figure 13b, depends on the value of A_2. Given that a complex linkage $i\,j, \ell$ is characterised by the dependency of w_{ij} on A_ℓ, if the weights of the FCM are deemed constant, complex linkages cannot be represented. Therefore, an FCM that has a constant weight matrix cannot include complex linkages.

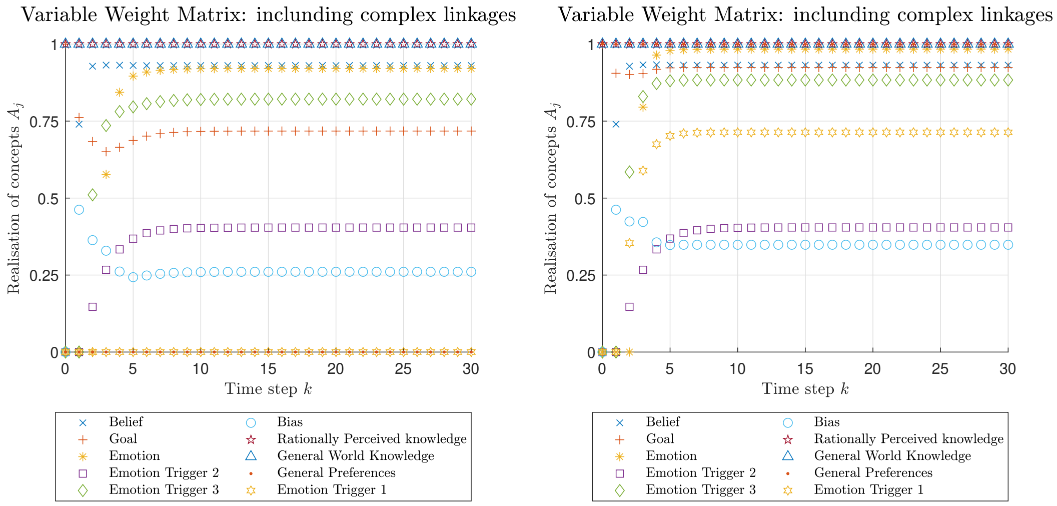
The effect of the belief of the agent over emotion trigger 3 increases/decreases as the goal increases/decreases. For example, if the agent really wants to do the outdoor activity and the weather is very good, emotion trigger 3 takes a higher value than if the agent does not have a goal and the weather is equally good. In fact, if the agent has a null goal of doing the activity, then the effect of the belief over emotion trigger 3 must be equally null. The influence of this complex linkage is visible when comparing two scenarios in which the belief held by the agent is similar, but the goal of the agent is different. In the simplest FCM (represented in Figure 13a), such a scenario is not easily represented. Even when the initial conditions are chosen to include the same initial rationally perceived knowledge for both scenarios and a different initial goal for each scenario, the goal quickly converges to a similar final value in both scenarios. This occurs because the simpler FCM has only one input, and that input takes the same (initial) value in both scenarios. Contrarily, the FCM that includes the slow-dynamics state variables has three inputs. Hence, with this FCM, it is possible to generate two scenarios in which the belief held by the agent is similar, but the goal is different. Two scenarios that fulfil this characteristic were selected. In both scenarios, the weather is very good, the accuracy of the weather forecast is very high and, initially, the agent had the goal to do the activity. In mathematical terms, this means that the rationally perceived knowledge, the general world knowledge and the initial goal of the agent were maximum ($A_7^{k=0} = 1$, $A_8^{k=0} = 1$ and $A_2^{k=0} = 1$). However, while in one scenario the general preference was null ($A_9^{k=0} = 0$), in the other it was maximum ($A_9^{k=0} = 1$). In linguistic terms, this means that in the first scenario the agents strongly likes the outdoor activity, while in the second the agent strongly dislikes it. The FCM was updated for 30 time steps for each one of the two scenarios. The realisations of the variables of the FCM after they converged were analysed.



(a) Evolution of the FCM variables over 30 time steps, in which the initial goal, the rationally perceived knowledge, and the general world knowledge are maximum ($A_2^{k=0} = 1$, $A_7^{k=0} = 1$, and $A_8^{k=0} = 1$), and the general preferences are null ($A_9^{k=0} = 0$).

(b) Evolution of the FCM variables over 30 time steps, in which the initial goal, the rationally perceived knowledge, the general world knowledge, and the general preferences are maximum ($A_2^{k=0} = 1$, $A_7^{k=0} = 1$, $A_8^{k=0} = 1$, and $A_9^{k=0} = 1$).

Fig. 16: Evolution of the FCM variables over 30 time steps, for two different scenarios, using an FCM in which the weights are constant.



(a) Evolution of the FCM variables over 30 time steps, in which the initial goal, the rationally perceived knowledge, and the general world knowledge are maximum ($A_2^{k=0} = 1$, $A_7^{k=0} = 1$, and $A_8^{k=0} = 1$), and the general preferences are null ($A_9^{k=0} = 0$).

(b) Evolution of the FCM variables over 30 time steps, in which the initial goal, the rationally perceived knowledge, the general world knowledge, and the general preferences are maximum ($A_2^{k=0} = 1$, $A_7^{k=0} = 1$, $A_8^{k=0} = 1$, and $A_9^{k=0} = 1$).

Fig. 17: Evolution of the FCM variables over 30 time steps, for two different scenarios, using an FCM that comprises complex linkages.

The first model (see Figure 16) yields the same final value of emotion trigger 3 for both scenarios, although the final value of the goal is lower in the first scenario (see Figure 16a) than in the second scenario (see Figure 16b). Contrarily, the second model yields a lower final value of emotion trigger 3 in the first scenario (Figure 17a), when compared to the second scenario (Figure 17b). This result is compatible with the fact that the goal is lower in the first scenario than in the second scenario.

These two examples show the importance of updating the weights of the FCM, both in terms of having simple linkages ij whose weights depend on the values of A_i and A_j, and in terms of including complex linkages. Given these two results that are obtained from a qualitative analysis of the FCM discrete time response, Hypothesis 1 is formulated. This hypothesis is tested in Section 5.3.2 using the data collected from the online survey with human participants.

Hypothesis 1 (H1): *Mathematically representing linkages as functions of the variables that are a part of the linkages, rather than representing them as constants (as done in the original FCM formulation), is essential to achieve accurate predictions of the beliefs, goals and emotions of rational agents.*

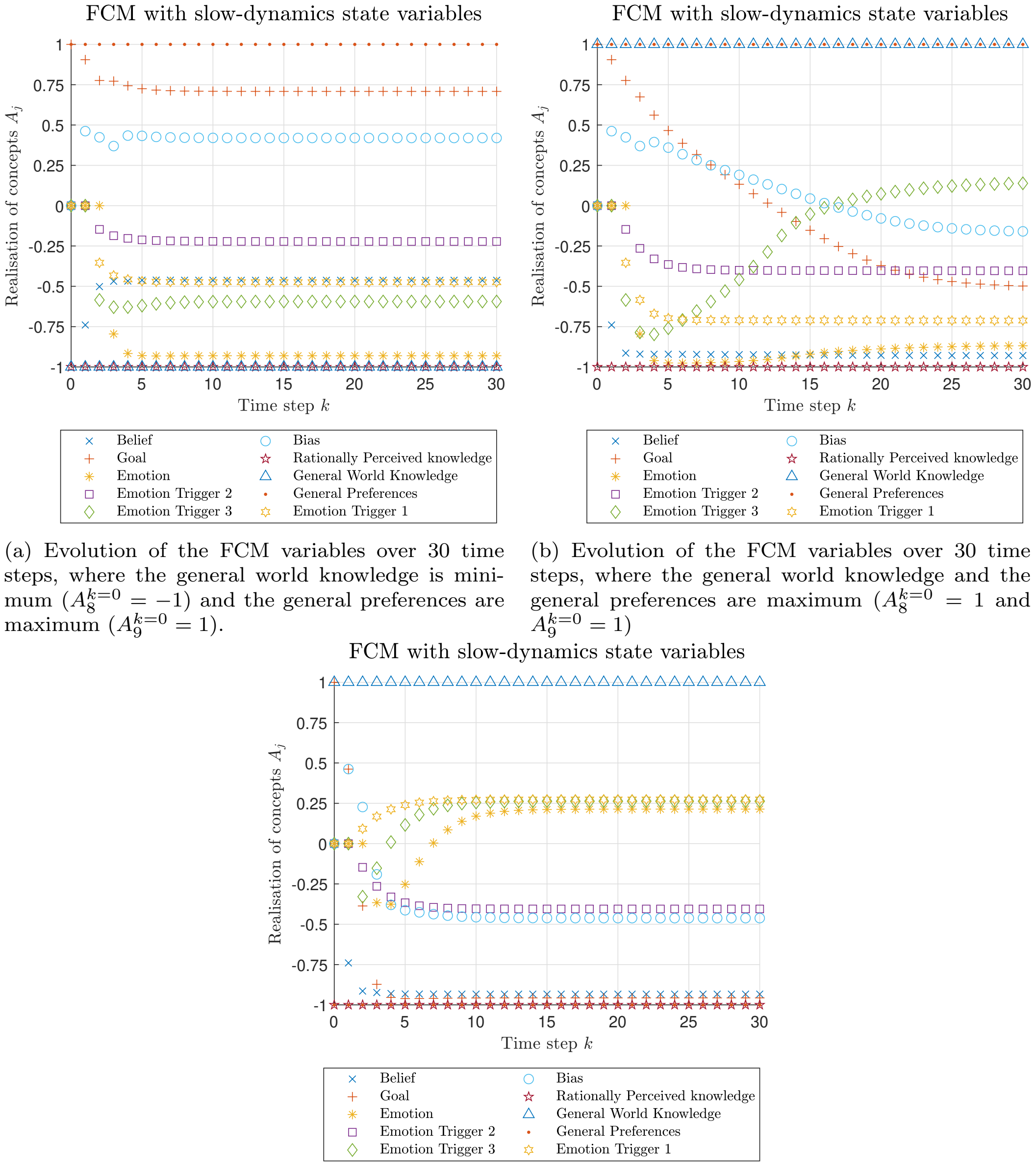
Furthermore, the analysis of the discrete time response of the variables of the FCM that represented the second real-life situation showed the importance of including the slow-dynamics state variables in the FCM. The importance of the general world knowledge can be seen by comparing two scenarios in which the general world knowledge is minimum in one and maximum in the other, while the values that define the other initial conditions are the same for both scenarios. For example, this comparison can be done considering two scenarios in which the rationally perceived knowledge is minimum ($A_7^{k=0} = -1$), and the initial goal and the general preferences are maximum ($A_2^{k=0} = 1$ and $A_9^{k=0} = 1$). This means that the weather is very bad, but the agent has an initial goal of doing the activity and strongly likes it. While in the first scenario the general world knowledge is maximum ($A_8^{k=0} = 1$), meaning that the agent knows that the weather forecast is usually accurate, in the second scenario the general world knowledge is minimum ($A_8^{k=0} = -1$), which implies that the agent knows that the accuracy of the weather forecast source is usually poor. The FCM that comprises the slow-dynamics state variables is able to capture the difference in the discrete time response of the belief of the agent when similar external conditions are applied, but a different general world knowledge is present. The belief of the agent is very similar to the rationally perceived knowledge when the trust on the weather forecast (general world knowledge) is maximum (see Figure 18b). However, when the trust on the weather forecast (general world knowledge) is minimum, the belief of the agent is not as high in magnitude as the rationally perceived knowledge, reflecting the higher uncertainty of the agent (see Figure 18a).

The importance of general preferences can be similarly verified. Two scenarios were selected to show the relevance of including the general preferences. In both scenarios, the initial goal of the agent and the general world knowledge are maximum, while the rationally perceived knowledge is minimum. These two scenarios differ in the initial value of the general preferences: while in one scenario the general preference is maximum, in the other scenario the general preference is minimum. In the scenario in which the general preference is minimum i.e., the agent strongly dislikes the outdoor activity, the final goal of the agent is close to the minimum (see Figure 18c). Intuitively, this result makes sense, since the weather is very bad and the agent strongly dislikes the activity. However, when the general preference is maximum i.e., the agent strongly likes the activity, the final goal of the agent is not as negative as in the previous case (see Figure 18b). In reality, the final goal of the agent is approximately $A_2^{k=30} - 0.5$, implying that the agent has medium intention of not wanting to do the activity, despite the very bad weather. Moreover, in the first scenario (Figure 18c), the final emotion of the agent takes a positive low value, implying that the agent feels slightly happy with the weather being bad. This emotion makes sense considering that the agent does not like the outdoor activity that it initially intended to do. In the second scenario (Figure 18b), the emotion takes a final strong negative value, which is a consequence of the external weather conditions not allowing the agent to do an activity that it strongly likes.

These two qualitative comparisons show the importance of including slow-dynamics state variables such as the general preferences and general world knowledge. Therefore, the following hypothesis was formulated:

Hypothesis 2 (H2): *Slow-dynamics state variables are essential to achieve accurate predictions of the beliefs, goals and emotions of rational agents.*

Finally, it is equally important to experimentally test the contribution of the personality traits in the model. As explained in the beginning of this section, the personality traits are mathematically represented in the FCM by choosing different weights for different individuals, rather than by a concept. Therefore, this variable allows the model to be personalised to different agents by tuning



(a) Evolution of the FCM variables over 30 time steps, where the general world knowledge is minimum ($A_8^{k=0} = -1$) and the general preferences are maximum ($A_9^{k=0} = 1$).

(b) Evolution of the FCM variables over 30 time steps, where the general world knowledge and the general preferences are maximum ($A_8^{k=0} = 1$ and $A_9^{k=0} = 1$)

(c) Evolution of the FCM variables over 30 time steps, where the general world knowledge is maximum ($A_8^{k=0} = 1$) and the general preferences are minimum ($A_9^{k=0} = -1$).

Fig. 18: Evolution of the FCM variables over 30 time steps in three time scenarios, using the FCM that comprises slow-dynamics state variables. In the three scenarios, the rationally perceived knowledge is minimum ($A_7^{k=0} = -1$), and the initial goal is maximum ($A_2^{k=0} = 1$). Therefore, these three scenarios correspond to situation where the weather is very bad, but the agent has an initial goal of doing the outdoor activity. The general world knowledge and general preferences differ for each scenario.

the weights to them, rather than by being represented by a concept that can take an initial value for each agent. Based on this premise and on the fact that the qualitative analyses that were performed in this section indicate that the other two slow-dynamics state variables must be included, a third hypothesis was formulated:

Hypothesis 3 (H3): *Personalising the weights of the FCM to each user is essential to achieve accurate predictions of the beliefs, goals and emotions of rational agents.*

Concept	Numerical value (Realisation of the concept)	Linguistic Term
General Preferences	-1	Dislike a great deal
	-0.66	Dislike a moderate amount
	-0.33	Dislike a little
	0	No preference
	0.33	Like a little
	0.66	Like a moderate amount
	1	Like a great deal
Rationally Perceived Knowledge	-1	Heavy Rain
	-0.5	Light Rain
	0	Unknown
	0.5	Cloudy
	1	Sunny
General World Knowledge	-0.4	Inaccurate
	0.2	Accurate
	0.8	Very accurate

Table 3: Possible values of the variables that defined the scenarios and linguistic terms associated to them. A scenario was defined by selecting a value in this table for the general preferences, rationally perceived knowledge, and general preferences.

Although the qualitative analysis of the discrete time response of the FCM did not directly lead to the formulation of Hypothesis 3, it is essential to test this hypothesis against the results of the survey.

5.3.2 Assessment of the model according to human participants

The assessment of the present model was performed by a quantitative comparison of the predictions of beliefs, goals and emotions made by the model implemented with an FCM, with the beliefs, goals and emotions of human participants. The beliefs, goals and emotions of human participants for different scenarios were collected through an online survey. Furthermore, the data collected in the survey was used to test the hypotheses formulated in Section 5.3.1.

Twenty-six scenarios were chosen to assess the model. The scenarios are defined by the initial values of three concepts: general world knowledge, general preferences and rationally perceived knowledge. In this section, these three concepts are referred to as *input concepts*. The scenarios were created selecting 1 out of 3 values for the general world knowledge, 1 out of 7 values for the general preferences, and 1 out of 5 values for the rationally perceived knowledge. Each one of these values was associated to a linguistic term and, in the survey, the initial conditions were given to the participants using those linguistic terms. Table 3 summarises the values that each one of the inputs could take in the scenarios, as well as the linguistic terms associated to each value.

The scenarios were divided into six sets. In the scenarios of a set, two of the three input concepts were constant, while the other one varied. In the first two sets, the rationally perceived knowledge was the input concept that varied from scenario to scenario. Each one of these two sets was composed by three scenarios. In the third and fourth sets, the general world knowledge was the variable input concept, and each one of these sets was composed by three scenarios. In the last two sets, the general preferences were the variable input, and each one of these sets was composed by seven scenarios (corresponding to all the possible values of the general preferences).

These twenty-six sets of scenarios were described to the participants during the online survey and, for each scenario, each participant was asked their belief, goal, and emotion. However, in order to produce questions that were intuitive to the participants, a different strategy from directly stating the initial conditions was adopted. For each scenario, requesting the user to picture a situation in which they wanted to do an outdoor activity that they liked a certain amount (for example, "dislike moderately") would require some reasoning effort from the participants and lead the answers to be less intuitive. Instead, the participants were initially allowed to categorise 17 predefined outdoors activities in terms of how much they liked each one of them, as shown in Figure 19a. Seven levels of preference were showed to the participants in linguistic terms. The only requirement of this ranking was that at least one activity was selected per level of preference. These seven linguistic terms, as well as the value of general preferences associated to each term,

Linguistic Term			**Numerical value**
Belief	**Goal**	**Emotion**	**(Realization of the concept)**
Heavy Rain	I do not want it at all	Very unhappy	-1
Light Rain	I do not want to do it	Unhappy	-0.5
I do not know	I have no preference	Nothing	0
Partially Sunny	I want to do it	Happy	0.5
Sunny	I want it a lot	Very happy	1

Table 4: Numerical values associated to the five linguistic terms placed above the sliding bars. These linguistic terms have the objective of making the process of answering the questions more intuitive and, consequently, ensuring more reliable results.

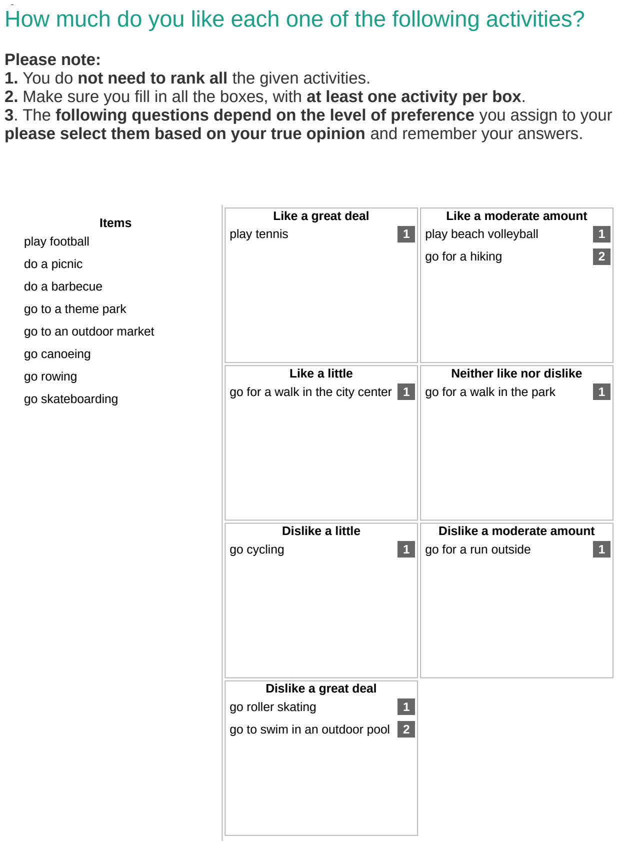
corresponded to the same linguistic terms and values that are presented in Table 3. Once the participants filled in all the levels of preference with at least one activity, the twenty-six scenarios were presented to them. As explained, each scenario was defined by the three values of the inputs. Thus, the scenarios were presented to the users with an introductory text that stated the linguistic terms corresponding to the values of the rationally perceived knowledge (the weather forecast) and of the general world knowledge (the accuracy of the source of weather forecast). Moreover, one of the activities that, during the initial question, was placed by the participant in the level of preference defined for the current scenario was randomly selected and placed in this introductory text. For example, Figure 19 shows the initial question where the participant ranks the activities in Figure 19a, and an example of a scenario proposed to the participant in Figure 19b. In this example, the participant placed the activity "go for a run outside" in the box "dislike a moderate amount" (see Figure 19a). Figure 19b shows a scenario that was displayed to the user, where the general preferences are "dislike a moderate amount", the rationally perceived knowledge is "partially sunny", and the general world knowledge is "moderately accurate". Therefore, to generate this question, one of the activities in the box "dislike a moderate amount" (in this case, "go for a run outside" is the only option) is shown to the participant. By presenting the information to the participants in this way, they do not have to reason as much to picture the scenario and can reply more intuitively.

After reading the introductory text of a scenario, the participants were inquired about their intention in the scenario (goal), how the situation made them feel (emotion) and what was their prediction regarding the outcome of the situation (belief). Figure 19b shows an example of the questions asked in a scenario. The participants were able to answer in a sliding scale that went from the linguistic term that corresponded to $A_j = -1$ to the linguistic term that corresponded to $A_j = 1$. Five linguistic terms were placed above the sliding scale, in order to allow the participants to answer intuitively. The realisations of the variables A_1, A_2, and A_3 that are associated to each linguistic term are summarised in Table 4.

Fifteen participants took part in this survey. However, only 14 of the answers were taken into account, given that the answers submitted by one of the participants were not correctly recorded.

An introductory text was given to the users in order to guide them and explain the structure of the survey. The participants were asked to picture the scenarios as well as they could, and they were informed that the questions asked in each scenario were relative to that scenario. Both pieces of information were repeated several times during the survey to ensure that the participants kept them in mind and that the answers were reliable.

To avoid errors caused by the lack of precision of the participants when moving the cursor, the answers to the sliding questions were discretised with a resolution of 0.1. Note that a resolution of 0.1 in the range $[-1, 1]$ does not have a negative impact on the results in the context of human cognition. The present model aims at predicting the cognitive states of rational agents. Those states are usually interpreted by people in linguistic terms, and the resolution of these linguistic terms is larger than the one considered in this discretisation. For example, for the state variable "emotion", the present discretisation admits 20 values between "very unhappy" and "very happy" (or 10 values between "no emotion" and "very happy"). However, when a person intuitively reasons about their emotions or the emotions of another person, they do not consider 20 levels of happiness. Instead, they attribute a smaller number of linguistic terms to describe the level of happiness.



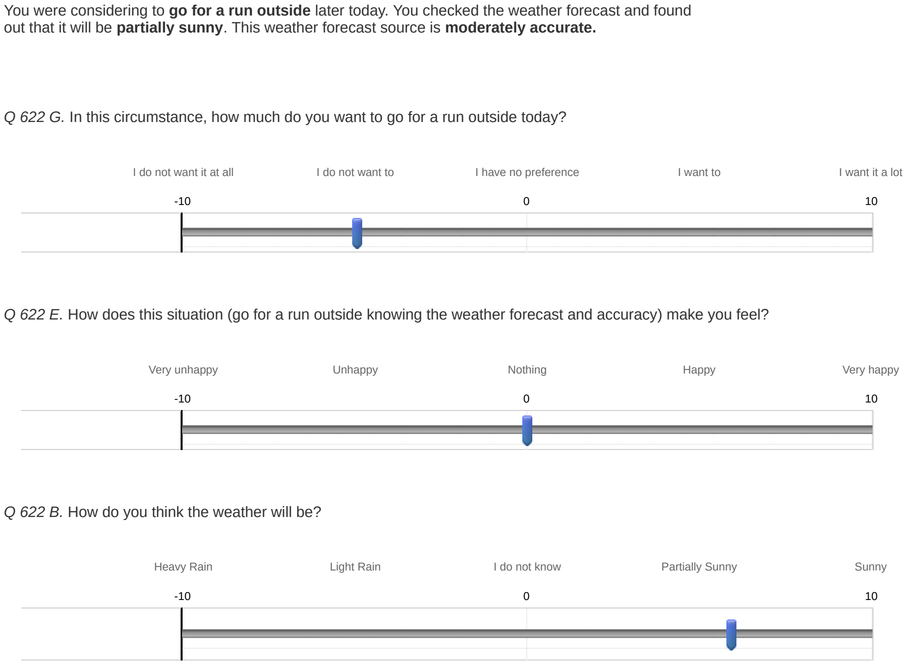
How much do you like each one of the following activities?

Please note:
1. You do **not need to rank all** the given activities.
2. Make sure you fill in all the boxes, with **at least one activity per box**.
3. The **following questions depend on the level of preference** you assign to your selected activities. So **please select them based on your true opinion** and remember your answers.

Items
play football
do a picnic
do a barbecue
go to a theme park
go to an outdoor market
go canoeing
go rowing
go skateboarding

Like a great deal	**Like a moderate amount**
play tennis 1	play beach volleyball 1 go for a hiking 2
Like a little	**Neither like nor dislike**
go for a walk in the city center 1	go for a walk in the park 1
Dislike a little	**Dislike a moderate amount**
go cycling 1	go for a run outside 1
Dislike a great deal	
go roller skating 1 go to swim in an outdoor pool 2	

(a) Initial question that collected the general preferences of the participants. For each one of the seven levels of general preference, the users were requested to assign at least 1 of the 17 outdoor activities to it. The participants were not required to rank all the activities. These question was only shown once to the user, at the beginning of the survey.



You were considering to **go for a run outside** later today. You checked the weather forecast and found out that it will be **partially sunny**. This weather forecast source is **moderately accurate.**

Q 622 G. In this circumstance, how much do you want to go for a run outside today?

Q 622 E. How does this situation (go for a run outside knowing the weather forecast and accuracy) make you feel?

Q 622 B. How do you think the weather will be?

(b) Example of a scenario that was shown to the participants, as well as the questions asked about that scenario. In the introductory text, the participants were indirectly informed of the initial conditions of that scenario. Subsequently, three questions that aimed at collecting the goal, emotion, and belief of the participants were asked. The participants could answer using the "slider" and five linguistic terms were placed above the scale of the slider to make the answers more intuitive. Twenty-five other scenarios similar to this one were shown to every participant.

Fig. 19: Structure of the questions asked in the online survey, and example of answers.

Subsequently, the output of the online survey was used to assess the model. In order to test the hypotheses stated in Section 5.3.1, the following four models were developed:

Model 1: FCM including all the model core, similarly to the model represented in Figure 13b. The linkages of the FCM were mathematically represented by functions. The model was personalised to each user.

Model 2: FCM *not including the slow-dynamics state variables nor emotion trigger 1*, similarly to the model represented in Figure 13a. The linkages of the FCM were mathematically represented by functions. The model was personalised to each user.

Model 3: FCM including all the model core, similarly to the model represented in Figure 13b. The linkages of the FCM were represented by *constant weights*. The model was personalised to each user.

Model 4: FCM including all the model core, similarly to the model represented in Figure 13b. The linkages of the FCM were mathematically represented by functions. The model was *not* personalised to each user.

As mentioned, the first three models were personalised to each user. This was done by personalising the weights or the functions associated to the linkages that were regulated by the personality traits in the model formalisation. Given that the personality traits regulated emotions, the functions associated to the linkages that influenced each one of the three emotion triggers were personalised (see Section 3.1.4). Similarly, since personality traits regulate the bias, the function associated to the linkage between the bias and beliefs was personalised (see Section 3.1.5). Finally, since personality traits regulate the intensity of the goals that are induced by beliefs, the function associated to the linkage between the belief and the goal was personalised (see Section 3.1.4).

As described in Section 5.3.1, representing some of these linkages by piecewise functions proved to be beneficial during the qualitative analysis. In these linkages, for each part of the domain that was previously defined by a sub-function, a weight was personalised. For example, the linkage between the belief and the goal (linkage 12) was defined by a piecewise function of the belief variable A_1, as described by Equation 5. In this case, two weights described the linkage. In similar cases, all the weights required to define the function that characterised the linkage were personalised.

In total, 9 weights were personalised for model 1, and 6 weights were personalised for both model 2 and model 3. For each one of these weights, 5 possible values were considered. The personalisation procedure of each model m consisted in a grid search that selected the combination of weights W_{comb} that, for each participant p, minimised the loss function $J_{p,m}$ given by Equation 6:

$$J_{p,m}(W_{\text{comb}}) = \sum_{s \in \mathbb{S}_t} \left(\sum_{j=\{1,2,3\}} \left(A_{j_m}(W_{\text{comb}}, s) - A_{j_p}(s) \right)^2 \right) \tag{6}$$

where $\mathbb{S}_t$ is the set of all scenarios considered in the training; $A_{j_m}(W_{\text{comb}}, s)$ is the final value (after convergence) of the realisation of concept C_j yielded by the FCM of model m when the initial conditions defined by scenario s are considered, and with the combination of weights W_{comb}; and $A_{j_p}(s)$ is the normalised value (in $[-1;1]$) that corresponds to the answer of the participant p regarding the concept C_j in scenario s.

The fourth model, which was not personalised, was still adapted to the population of participants. The 9 weights that were considered to be customisable in model 1 were adapted in model 4. More specifically, the combination of these weights W_{comb} that minimised the loss function $J_{\text{model 4}}$, which is given by the following equation, was selected:

$$J_{\text{model 4}}(W_{\text{comb}}) = \sum_{p \in \mathbb{P}} \left[\sum_{s \in \mathbb{S}_t} \left(\sum_{j=\{1,2,3\}} \left(A_{j_{\text{model 4}}}(W_{\text{comb}}, s) - A_{j_p}(s) \right)^2 \right) \right] \tag{7}$$

where p is the participant and $\mathbb{P}$ is the set of all participants.

The difference between model 1 and model 4 is that, while in model 1 an optimal combination of weights was found for each participant, in model 4 only one combination of weights was selected for all the participants.

Finally, it is important to explain that not all the scenarios were used to personalise the models, since around 30% of the data should be saved to assess them. Nonetheless, the amount of data available for training and testing (26 scenarios) is limited. In a first analysis, the first five data sets were used for training and the last data set, which was composed of 7 scenarios, was used for testing. However, 7 scenarios were not enough to assess the models. Increasing the length of the survey was not a viable option, since the participants took 33 minutes on average to complete it. Making it larger would increase the risk of the participants getting tired or distracted and, consequently, providing inaccurate answers.

Therefore, to increase the amount of testing data without hindering the quality of the results, each model was trained in three independent runs, using different batches of scenarios to train and test the models in each run. In the first run, the models were trained using the third through sixth data sets, and were tested using the first two data sets. Thus, six scenarios were used in the first run to test the models. In the second run, the first two and last two data sets were used to train the models. The assessment was performed using the third and fourth data sets, which had in total six scenarios. Finally, in the third run, the first five data sets were used to train the models, and the last data set, which had seven scenarios, to test them. By following this procedure, the results were made more reliable.

The models are assessed based on the comparison between the value of each fast-dynamics state variable A_1, A_2, and A_3 given by each model m in each scenario s of the testing data $\mathbb{T}$, with the answers given by the participants about each fast-dynamics state variable A_1, A_2, and A_3, for each scenario s of the testing data $\mathbb{T}$. More specifically, the average of the square of the difference between the prediction and the true value over all the participants is computed. This mean squared error is defined for each fast-dynamics state variable A_j, each scenario s of the testing data $\mathbb{T}$ and each model m, as described by the following equation:

$$\text{MSE}(j,s,m) = \frac{1}{P}\sum_{p\in\mathbb{P}}(A_{j_m}(W^*_{\text{comb}}(m,p),s) - A_{j_p}(s))^2 \ , \qquad \text{for } j = \{1,2,3\},\ m = \{1,2,3,4\},\ s \in \mathbb{T} \quad (8)$$

where W^*_{comb} is the optimal combination of weights found for model m and participant p in the optimisation process, P is the number of participants and $\mathbb{P}$ is the set of all participants. The data was tested using absolute errors, given that the outputs were normalised in the interval $[-1;1]$. The average error of a model m over a testing set $\mathbb{T}_i$ in the prediction of a variable j is given by:

$$\text{MSE}_{\mathbb{T}_i}(j,m) = \frac{1}{N_i}\sum_{s\in\mathbb{T}_i}\left(\frac{1}{P}\sum_{p\in\mathbb{P}}(A_{j_m}(W^*_{\text{comb}}(m,p),s) - A_{j_p}(s))^2\right) , \qquad \text{for } j = \{1,2,3\},\ m = \{1,2,3,4\} \quad (9)$$

where N_i is the number of scenarios in the testing set $\mathbb{T}_i$.

5.4 Discussion of the results

The results obtained in the assessment of the model are summarised in Table 5, Table 6 and Table 7. Table 5 presents the mean squared error in the prediction of beliefs that was attained by every model over each testing set, as well as in general. Table 6 shows the mean squared error in the prediction of goals achieved by each model over each testing set, and in general. Table 7 presents the mean squared error in the prediction of emotions shown by each model over each testing set, and in general. Since the range of the values is $[-1,1]$ and the errors are squared, it is important to take into consideration that the maximum possible error is 4.

When comparing the mean squared error obtained by models 1 and 2, the first model achieves a lower error in all the runs for the three fast-dynamics state variables. Overall, the mean squared error considering all the testing data achieved by the first model is lower than the one achieved by the second model, for the three fast-dynamics state variables. Therefore, Hypothesis 2 is supported by the experimental data: slow-dynamics state variables are relevant to achieve accurate results in predicting the cognitive concepts of rational agents.

	First run	Second run	Third run	All the testing data
Model 1	0.473	0.148	0.057	0.218
Model 2	0.541	0.234	0.059	0.264
Model 3	0.493	0.393	0.085	0.314
Model 4	0.549	0.156	0.105	0.261

Table 5: Mean squared error in belief prediction for each model. For every run, a different testing set was considered. The final column corresponds to the average error over all the scenarios that were used for testing.

	First run	Second run	Third run	All the testing data
Model 1	0.386	0.400	0.115	0.314
Model 2	0.426	0.793	0.133	0.461
Model 3	0.548	0.744	0.418	0.593
Model 4	0.373	0.406	0.219	0.353

Table 6: Mean squared error in goal prediction for each model. For every run, a different testing set was considered. The final column corresponds to the average error over all the scenarios that were used for testing.

	First run	Second run	Third run	All the testing data
Model 1	0.185	0.198	0.120	0.162
Model 2	0.255	0.268	0.183	0.234
Model 3	0.244	0.288	0.141	0.223
Model 4	0.211	0.257	0.128	0.195

Table 7: Mean squared error in emotion prediction for each model. For every run, a different testing set was considered. The final column corresponds to the average error over all the scenarios that were used for testing.

When comparing the performance of the first and third models, model 1 achieves a substantially lower mean squared error in all the runs, for all the variables. Moreover, the overall error achieved by the model 1 is significantly lower than the one achieved by model 3, specially in the prediction of goals. These results support Hypothesis 1 i.e., linkages should be represented by functions.

Finally, by comparing the mean squared error obtain by model 1 against the one obtained by model 4, it can be concluded that model 1 made more accurate predictions of beliefs and emotions than model 4, in all the runs. Compared to mode 4, the error of model 1 in predicting the goals of humans was 0.013 higher in the first run, 0.006 lower in the second run, and 0.104 lower in the third run. Therefore, the error in goal prediction was similar for both models in the first two runs, and lower for model 1 in the third run. Overall, the error in goal prediction of model 1 was lower than the one achieved by model 4. These results uphold Hypothesis 3, showing that the weights of the FCM must be personalised to each individual and that personalisation is important.

Furthermore, considering all the testing data, model 4 has the best overall performance of models 2, 3 and 4. This result shows that a global model is a good starting point when dealing with an unknown individual, which can be later on personalised to that individual once the data that allows the personalisation to be carried is available (i.e., once more interactions are carried out).

6 Conclusions

The present research project formalises, formulates, implements, and tests a cognitive model that aims to address the main challenges currently faced by Socially Assistive Robots (SARs). An extensive literature survey on SARs shows that SARs lack the capability of retaining the attention of their users for long periods of time and that a comprehensive model of the cognitive processes that drive human behaviour has the potential to tackle this issue (Section 1).

During the formalisation of the model in Section 3, two classes of state variables are identified. One of these classes consists in fast-dynamics state variables. These variables influence the behaviour of the agent in every interaction and are modified with high frequency. Three fast-dynamics state variables are considered: beliefs, goals and emotions. Beliefs correspond to short-term knowledge about the world as perceived by the agent, and this knowledge does not have to correspond to reality. Goals correspond to the short-term intentions and desires of the agent. Emotions describe the emotional state of the agent. Secondly, the model comprises slow-dynamics state variables, which are low-frequency variables that shape the behaviour of the agent throughout interactions. These low-frequency variables, which can be interpreted as parameters, are specific to each rational agent and, for that reason, allow personalising the model. Three slow-dynamics state variables are identified: general world knowledge, general preferences, and personality traits. The general world knowledge consists in the permanent knowledge that an agent has acquired about the world that surrounds it, and directly affects the beliefs of the agent. The general preferences correspond to the tastes and interests of the rational agent, and mainly influence its goals. Finally, the personality traits correspond to the defining aspects of the personality of the agent, and regulate how most of the variables interact with each other.

The process of belief development is formalised as two separate sub-processes, as described in Section 3.1.3. The first corresponds to the act of perceiving data from the information available in the real world. The output of this process corresponds to raw, unprocessed data that is internal to the agent. This perceived data is then processed by the agent in a procedure named rational reasoning, generating the rationally perceived knowledge. Finally, Section 3.1.5 describes how emotions and goals are found to influence the process of constructing beliefs. This influence, which is regulated by the personality traits of the rational agent, is represented by a bias variable that, jointly with the rationally perceived knowledge, generates the perceived knowledge.

Prior research had focused on how beliefs and goals interact with each other and how they trigger actions. Nonetheless, the analysed ToM models had not comprised emotions so far. Thus, the role of emotions in the model is thoroughly discussed in Section 3.1.4. It is concluded that emotions are triggered either by beliefs, combinations of beliefs and goals, and combinations of beliefs and general preferences. Moreover, they are regulated by personality traits.

Section 4 formulates the model. To simplify this procedure, the model is divided in five modules that can be formulated independently. The formulation is focused on the model core, which corresponds to the internal state variables that are not involved in the observation and reasoning processes, as well as their relationships. Two formulations are proposed: one based on probability theory, and the other on fuzzy logic. In the first, the model core is formulated as a Bayesian Network (BN), while the input and output processes of the model core, as well as the world model, are formulated as a Partially Observable Markov Decision Process (POMDP). Nonetheless, implementing this formulation would require altering the structure of the formalised model. The second formulation suggests representing the model core as a Fuzzy Cognitive Map (FCM), and the input and output processes of the model core with a proper decision-making framework. Given the more intuitive nature of FCMs and the possibility to main the original structure of the formalised model, the second formulation is chosen for the implementation.

The model implementation is explained in Section 5. To implement and test the model, a real-life situation is defined. Based on the selected situation, the model is represented as an FCM and the resultant map is implemented in MATLAB. First, the discrete time response of the model to different situations is qualitatively analysed. Based on this analysis, three hypotheses are formulated regarding the importance of representing the FCM linkages as functions, of slow-dynamics state variables, and of personalising the functions that represent the linkages to each user. Finally, an online survey with 15 participants is carried out to collect data to test the model. Twenty-six initial conditions of the real-life situation were presented to the participants during the survey. For each scenario, the participants evaluated what corresponded to their beliefs, goals and emotions. To test the three hypotheses, four models are developed. The first model comprises the entire model core, and represents the FCM linkages through non-constant functions. The second model differs from the first by not including the slow-dynamics state variables. The third model is similar to the first, apart from the fact that the linkages are represented through constant weights. The functions that represent the linkages of these three models are personalised to every user. Finally, the fourth model is similar to the first model, but it is not personalised to each user. Instead, the weight functions are selected to best fit the entire population. The performance of the models is evaluated based on the mean squared error between the predictions made by each model with the data provided by the participants. By comparing the performance of the first model with each one of the other three, it was possible to conclude that both including the slow-dynamics state variables and representing the linkages of the FCM as functions is essential to achieve accurate predictions. Moreover, personalising the functions that represent the linkages to each user yields more accurate predictions than using a general model (the fourth model). However, the results of the fourth model are satisfactory, which indicates that this model can be used as a starting point with an unknown agent and later on personalised based on the data collected throughout interactions.

6.1 Directions for Future Work

In the future, the developed model will be applied to more complex situations. Ideally, we want to attain a model that can predict the cognitive states of rational agents in general situations. Moreover, we want to apply the model as a predictive module in the loop of the control system that generates the behaviour of the SAR.

For this reason, the FCM should be extended to account for multiple concepts that are associated to the same model variable. For example, the fast-dynamics state variable emotions was represented in the FCM by one FCM concept that reflected the level of happiness. Nonetheless, there are other emotions that influence the human cognition and behaviour, and should be represented in the FCM. Consequently, it is important to develop a framework that enables the usage of the general structure of the formulated model, while supporting that each model variable has several FCM concepts associated to it. This is essential to allow the same model to deal with multiple, more complex situations.

To do so, several measures have to be taken to increase the computational efficiency of the FCM. So far, the FCM was implemented using MATLAB. In the context of the present research, time efficiency was not a hard constraint, given that all the experiments were run offline. The personalisation of one model, which was the most time-consuming procedure, was in the order of minutes. Moreover, executing the model to estimate a belief-goal-emotion triplet was in the order of the milliseconds. If the present model is used to estimate the cognitive mental states of rational agents in predictive control, an acceptable sampling frequency of the controller is around 1 second. In this case, the computational time taken by the current model would not compromise the fulfilment of this requirement. Therefore, for the time being, there was no need to improve the computational efficiency. Nevertheless, if the level of complexity of this model is raised, the time that is needed to personalise the model (in the training phase) will substantial raise, as well as the necessary time to obtain a prediction from the model during execution. When the model is used to predict the cognitive states of an user during an interaction, it is essential that the FCM is implemented in such a way that its execution speed does not compromise the interaction. To address these potential issues, the usage of a lower level programming language, such as python or C++, is recommended. Furthermore, to accelerate the personalisation process, an optimisation algorithm that is more advanced than grid search (such as policy gradient) is likely to be required.

Finally, during the present research project, there was no possibility to carry out in-person interactions between humans and a SAR. Although the present results show the ability of this model to predict human cognitive states, in order to assess the impact that the developed model can have on the social skills showed by SARs, these interactions must be performed. Moreover, it is important to analyse how much this model can improve the engagement of users in SARs during long-term interactions, which can only be done if interactions that are composed of several sessions are carried out.

References

Andrade EB, Ariely D (2009) The enduring impact of transient emotions on decision making. Organizational Behavior and Human Decision Processes 109(1):1–8

Baker CL (2012) Bayesian Theory of Mind : modeling human reasoning about beliefs, desires, goals, and social relations. PhD thesis, Massachusetts Institute of Technology

Baker CL, Tenenbaum JB, Saxe RR (2007) Goal Inference as Inverse Planning. In: Proceedings of the Annual Meeting of the Cognitive Science Society, Nashville, TN, USA, pp 779–784

Baker CL, Goodman ND, Tenenbaum JB (2008) Theory-based Social Goal Inference. In: Proceedings of the Annual Meeting of the Cognitive Science Society, Washington DC, USA, pp 1447–1452

Baker CL, Saxe RR, Tenenbaum JB (2011) Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. In: Proceedings of the Annual Meeting of the Cognitive Science Society, Boston, MA, USA, pp 2469–2474

Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB (2017) Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nature Human Behaviour 1(4):1–10

Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a "theory of mind" ? Cognition 21(1):37–46

Bono JE, Vey MA (2007) Personality and emotional performance: Extraversion, neuroticism, and self-monitoring. Journal of Occupational Health Psychology 12(2):177–192

Carvalho JP, Tomé JA (2001) Rule based fuzzy cognitive maps - Expressing time in qualitative system dynamics. In: IEEE International Conference on Fuzzy Systems, Melbourne, VIC, Australia, pp 280–283

Cheah WP, Kim KY, Yang HJ, Kim SH, Kim JS (2008) Fuzzy Cognitive Map and Bayesian Belief Network for Causal Knowledge Engineering: A Comparative Study. The KIPS Transactions: Part B 15B(2):147–158

Clabaugh C, Mahajan K, Jain S, Pakkar R, Becerra D, Shi Z, Deng E, Lee R, Ragusa G, Matarić M (2019) Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. Frontiers in Robotics and AI 6(110):1–18

Dennett DC (1987) The Intentional Stance. MIT Press, Cambridge, MA, USA

Feil-Seifer D, Matarić MJ (2005) Defining socially assistive robotics. In: Proceedings of the 9th International Conference on Rehabilitation Robotics, Chicago, IL, USA, pp 465–468

Foka A, Trahanias P (2007) Real-time hierarchical POMDPs for autonomous robot navigation. Robotics and Autonomous Systems 55(7):561–571

George JM, Dane E (2016) Affect, emotion, and decision making. Organizational Behavior and Human Decision Processes 136:47–55

Heckerman D (2008) A tutorial on learning with Bayesian networks. In: Innovations in Bayesian Networks: Theory and Applications, vol 156, 1st edn, Springer, Berlin, Heidelberg, Berlin, Heidelberg, Germany, pp 33–82

Jara-Ettinger J, Gweon H, Schulz LE, Tenenbaum JB (2016) The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. Trends in Cognitive Sciences 20(8):589–604

Johnson-Laird PN (1994) Mental models and probabilistic thinking. Cognition 50(1-3):189–209

Kaelbling LP, Littman ML, Cassandra AR (1998) Planning and acting in partially observable stochastic domains. Artificial Intelligence 101(1-2):99–134

Kahraman C, Deveci M, Boltürk E, Türk S (2020) Fuzzy controlled humanoid robots: A literature review. Robotics and Autonomous Systems 134:1–12

Kidd CD, Breazeal C (2008) Robots at home: Understanding long-term human-robot interaction. In: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Nice, France, pp 3230–3235

Kwon DS, Chung MJ, Park JC, Yoo CD, Jee ES, Park KS, Kim YM, Kim HR, Park JC, Min HJ, Park JW, Yun S, Lee KW (2008) Emotional Exchange of a Socially Interactive Robot. In: 17th IFAC World Congress, Seoul, Korea, pp 4330–4335

Lee JJ, Sha F, Breazeal C (2019) A Bayesian Theory of Mind Approach to Nonverbal Communication. In: ACM/IEEE International Conference on Human-Robot Interaction, Daegu, Korea, pp 487–496

Leite I, Martinho C, Paiva A (2013) Social Robots for Long-Term Interaction: A Survey. International Journal of Social Robotics 5(2):291–308

Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. Annual Review of Psychology 66:799–823

Matarić MJ, Scassellati B (2016) Socially assistive robotics. In: Springer Handbook of Robotics, 1st edn, Springer, Berlin, Heidelberg, Berlin, Heidelberg, Germany, pp 1973–1993

Mourhir A, Rachidi T, Papageorgiou EI, Karim M, Alaoui FS (2016) A cognitive map framework to support integrated environmental assessment. Environmental Modelling & Software 77:81–94

Rabinowitz NC, Perbet F, Song HF, Zhang C, Eslami SM, Botvinick M (2018) Machine theory of mind. In: Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, pp 4218–4227

Raghunathan R, Pham MT (1999) All negative moods are not equal: Motivational influences of anxiety and sadnesson decision making. Organizational Behavior and Human Decision Processes 79(1):56–77

Saxe R, Houlihan SD (2017) Formalizing emotion concepts within a Bayesian model of theory of mind. Current Opinion in Psychology 17:15–21

Scassellati B (2002) Theory of Mind for a Humanoid Robot. Autonomous Robots 12(1):13–24

Scassellati B, Boccanfuso L, Huang CM, Mademtzi M, Qin M, Salomons N, Ventola P, Shic F (2018) Improving social skills in children with ASD using a long-term, in-home social robot. Science Robotics 3(21):1 –9

Sedki K, Bonneau de Beaufort L (2012) Cognitive Maps and Bayesian Networks for Knowledge Representation and Reasoning. In: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, Athens, Greece, pp 1035–1040

Stylios CD, Groumpos PP (2004) Modeling Complex Systems Using Fuzzy Cognitive Maps. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 34(1):155–162

Tapus A, Mataric MJ (2008) Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In: AAAI Spring Symposium, Stanford, CA, USA, pp 133–140

Wang Q, Jeon HJ (2020) Bias in bias recognition: People view others but not themselves as biased by preexisting beliefs and social stigmas. PLOS ONE 15(10):1–18

Wellman HM, Cross D, Watson J (2001) Meta-analysis of theory-of-mind development: The truth about false belief. Child Development 72(3):655–684

Zaki J, Craig Williams W (2013) Interpersonal emotion regulation. Emotion 13(5):803–810

II

Literature Study

3 Introduction

Robots assist humans in a multitude of tasks. However, while most robots are employed in industrial manufacturing tasks, there is a class of robots whose objective is to aid individuals in personal tasks through human-robot social interactions (Clabaugh & Matarić, 2018). These robots, known as Socially Assistive Robots (SARs), aim at improving their users' quality of life in fields such as health care (Kidd & Breazeal, 2008; Tapus & Mataric, 2008), education (Clabaugh et al., 2019), and daily life companionship and chores assistance (Rodríguez-Lera et al., 2018).

These robots have shown the potential to improve human's quality of life (Clabaugh et al., 2019; Scassellati et al., 2018) and, ultimately, are expected to help users on a regular basis in their private sphere, for an indefinite period of time (Matarić & Scassellati, 2016). Nevertheless, SARs are not yet capable of retaining users' attention for long periods of time, since they lack the social skills that would enable them to create rich and natural interactions (Leite et al., 2013). This lack of social skills results from an absence of deep understanding of the perceptions, cognition, and state-of-mind of the human in response to various stimuli. Therefore, these robots are still not advanced enough to aid their users in various real-world environments and to keep them actively engaged in human-SAR interactions (Leite et al., 2013). Moreover, only a few of the current SARs are commercially available (Clabaugh & Matarić, 2018).

Given the current challenges faced by socially assistive robotics and the potential presented by this field to aid people from different social groups and needs (Matarić & Scassellati, 2016), a literature survey was conducted to investigate the current paradigm of the control approaches that steer decisions and actions of SARs, to detect current flaws and research gaps, and to identify new potential control approaches for SARs that improve their performance and result in more engaging and lasting interactions with humans. Furthermore, it is known that social connections and cooperation between a user and a SAR can boost the effectiveness of the SAR (Clabaugh & Matarić, 2018; Tapus & Mataric, 2008). As a result of this, the field of SARs relies not only on engineering and computer science knowledge, but also on social sciences such as psychology (Clabaugh & Matarić, 2018). Therefore, while the main focus of this survey is on the engineering and, especially, systems and control discipline, some relevant aspects regarding social sciences that are necessary to understand the interactions between humans and SARs were also discussed.

Currently, the control techniques that are employed on SARs mainly focus on displaying complex behaviours and skills, rather than acknowledging and perceiving the characteristics and mental states of the human (Leite et al., 2013; Rodríguez-Lera et al., 2018). Given the importance of mutual understanding in social interactions, developing new approaches to model the cognition of users can address the current challenges that SARs face, especially regarding sustaining long-term engaging interactions with humans (Matarić & Scassellati, 2016), and thus will contribute to bringing socially assistive robotics closer to its final goal.

For that reason, the second part of this survey analyses the state-of-the-art models that represent human cognition and behaviour, as well as other potential approaches that can be employed to develop such a model. Given the success of prior models in explaining certain parts of human cognition (Baker et al., 2011; Baker et al., 2007), by investigating the characteristics of these models while keeping in mind the remaining challenges in human-SARs interactions, a method capable of tackling these challenges is expected to be identified. Subsequently, in the execution of this thesis, the identified approaches will be integrated, adapted and/or improved to develop a model of the user's cognition. This model should estimate the current state-of-mind of the user based on their observable behaviours and should predict the future state-of-mind of the user as a result of the current stimuli the user is dealing with. All in all, such model is expected to provide SARs with the means to behave realistically and naturally in social interactions by properly accompanying users and understanding their progress throughout long-term interactions.

The following sections of this chapter formalise the objectives of the present research project, the methodologies and scientific tools selected to attain such objectives, and the questions that this study will aim to answer. Moreover, the overview of the remaining literature survey is presented.

3.1. Research Objectives

The overarching research objective of this thesis is:

> "To contribute to the development of more realistic and engaging human-SAR long-term interactions by means of formalising and mathematically modelling the cognition and decision making processes of humans within the Theory of Mind framework".

In order to achieve the aforementioned objective, this research will focus on attaining sub-objectives that contribute to the fulfilment of the main objective. Firstly, it is important to identify the elements that contribute to the cognition of humans and to the human-human interactions and the connections and influences of these elements. Based on the achievements of the first sub goal, we will develop a model that describes the cognition and decision-making process of humans as a function of the identified elements. Subsequently, it is necessary to mathematically formulate the proposed model. Finally, the model will be implemented, tested and validated by comparing the estimations made by the model with those made by humans.

3.2. Research Questions

This subsection describes the three main questions that the present study will provide answers to. The answer to these questions will allow to fulfil the research objectives. Nonetheless, in order to answer these questions, simpler sub-questions, for which the answers contribute to provide an answer to the higher level questions, must be addressed.

For reference, the mathematical model of human's cognition will include variables that correspond to mental states of humans (e.g., goals, desires, emotions), and will present the processes that develop, regulate, or update these states. The state variables of the model belong to two categories:

1. State variables that can be identified via long-term interactions and that correspond to the traits that are inherent to individuals and - in this research - are considered constant[1] throughout human-SAR interaction sessions;

2. State variables that are generated in or evolved short-term interactions and with a much higher frequency than the first category of state variables, and correspond to transient mental states of humans.

The research questions, as well as the sub-questions that should be answered in order to find proper answers for the main research questions, are the following:

- **Question 1:** What knowledge and structure should a cognitive model possess to be able to accurately describe the perception, cognition, and reasoning processes behind the decisions and behaviour of humans when they are engaged in long-term human interactions? More specifically, the following sub-questions will be addressed:
    - Which internal elements (state variables and processes) are common in and relevant to human interactions?

[1] To be more accurate, in this research these state variables are treated as parameters, but in reality they can evolve in time with a low frequency.

  - Which elements influence or are influenced by others?
  - Which elements affect or are affected by short-term interactions and which affect or are affected by long-term interactions?
  - What processes are involved/can be identified when internal variables of a cognitive model interact with the external real world?

- **Question 2:** What are the most adequate approaches to formalise and formulate the theoretical model of human's cognition? To answer this question, the following sub-questions will be investigated:
  - Which approaches from the state-of-the-art can propose a structure that properly fits the requirements of the proposed cognitive model?
  - How do the available approaches compare in terms of accuracy and computational efficiency?
  - What are the open challenges in the state-of-the-art approaches that should be tackled?

- **Question 3:** To what extent can the developed model accurately explain or predict the perceptions and decision-making procedures of humans in long-term interactions? In particular, the subsequent questions will be studied:
  - What is the accuracy of the estimations made by the model about the mental state of a user compared to the feedback that will directly be provided by the users about their real mental states?
  - To what extent do the variables that influence long-term interactions increase the model's accuracy compared to a cognitive model that considers only those variables that influence the short-term interactions?
  - Does the accuracy of the model's predictions improve throughout sessions?

Although a comparison between the effectiveness of a controller developed based on this model with the previous model-free controllers found in literature is relevant, considering the time available it will not be feasible to consider this comparison as a part of this research. For that reason, further investigation corresponding to this research question is included in our recommendations for future research.

3.3. Report Structure

The remained of this survey consists of two main chapters and the conclusions.

In Chapter 4, an overview of the state-of-the-art human-SAR interactions is presented. This chapter includes a description of SARs in Section 4.1, an analysis of the nature of interactions between humans and SARs in Section 4.2, and a description of the methods and control techniques employed in the state-of-the-art socially assistive robotics in Section 4.3. Additionally, the differences in length of the human-SAR interactions, and the challenges faced by both studies that comprise short-term interactions and studies that consider long-term interactions are described in Section 4.4. Section 4.5 concludes Chapter 4, presenting an overview of the main challenges and pitfalls detected in the current human-SAR interactions, as well as a potential approaches to overcome these challenges.

Subsequently, Chapter 5 focuses on the approaches used in the literature to model human cognition and decision making processes. The first two chapters of this section describe the Theory of Mind (ToM), a framework that explains how rational agents, including humans, perceive and reason about the mental states of other rational agents. Although this framework was initially proposed by the neuroscience community, it was later adapted by researchers in other disciplines, e.g., robotics and computer science, in order to develop a virtual ToM that can be applied in machines to better understand human reasoning and interact with humans. Thus, while Section 5.1 considers the main concepts of ToM from a neuroscience point of view, Section 5.2 focuses on the Bayesian Theory of Mind (BToM), an approach that employs Bayesian inference to represent the ToM and that is widely used in literature. Section 5.3 describes the concept of meta-learning and how this concept has been used in prior research to develop a virtual ToM. Section 5.4 describes the theory behind Markov Decision Process (MDP) and Partially Observable Markov Decision Process (POMDP), and the projects where this framework has been employed to model human's cognition. Finally, Section 5.5 presents the most important outcomes of Chapter 5.

The detected research gaps, the methods proposed to tackle these gaps and the future directions of the present research project are described in Chapter 6, which finalises this literature survey.

4

Decision making in human-SAR interactions

The present chapter reviews and analyses the controllers that steer the decision and actions of SARs when interacting with humans. We will focus on the control techniques that are applied to SARs in the literature, in order to identify the main weaknesses of these approaches and to come up with proposals to overcome these shortcomings.

In Section 4.1, an official definition of SARs is presented. Moreover, the fields in which these robots assist humans and the most important characteristics that SARs must possess are discussed.

Section 4.2 describes the interactions between humans and SARs in different fields of application of these robots. This description encompasses the goals of the analysed research projects, the tasks in which the robots help their users, the role of the SARs in each task or project, and a description of the robots used or designed.

Section 4.3 covers the main control methods employed in SARs found in literature. This section was based both on the techniques applied by the analysed research projects as well as other control techniques commonly used in the field of SARs, as described by other literature surveys. Moreover, some authors resorted to the implementation of certain concepts or techniques which cannot be considered control approaches but work in tandem with the controller and are important to shape the robot's decision making. In the cases in which these techniques or concepts were relevant or essential for the controller, they are discussed in this section.

Finally, Section 4.4 classifies the studies analysed in the previous sections according to their duration, i.e., whether they are considered short-term or long-term interactions. This distinction is of utmost importance in human-SAR interactions since the resources employed and the conclusions that can be taken from each one of these two categories are different. Given the additional challenges inherent to long-term interactions and the focus of the present project on this type of interactions, most of the studies analysed are comprised in this category. Consequently, Section 4.4 analyses long-term interactions more extensively than short-term ones.

4.1. Socially Assistive Robots

According to Feil-Seifer and Matarić (2005), socially assistive robotics can be described as the merge of two robotic fields that had earlier appeared: assistive robotics and socially interactive robotics.

On the one hand, assistive robotics refers to a robotic field that concerns the development of robots that provide physical assistance mainly for people with physical disabilities. Nonetheless, given that this definition leaves out those Assistive Robots (ARs) that help users by employing non-physical contact, a more updated definition was proposed by the same authors, characterising an AR as "one that gives aid or support to a human user", regardless of the type (i.e., physical or non-physical) of this help (Feil-Seifer & Matarić, 2005).

On the other hand, socially interactive robotics encompasses robots that autonomously interact with users in a social context, excluding robots that are remotely controlled by humans. As opposed to

ARs, Socially Interactive Robots (SIRs) usually aim at entertaining the user rather than assisting them (Feil-Seifer & Matarić, 2005).

Nevertheless, prior studies have shown that, in some fields or situations such as rehabilitation and special needs education, engaging in social interactions with humans while helping them with their rehabilitation or educational exercises yields the most effective results in terms of accomplishment of therapeutic physical tasks and learning outcome (Clabaugh et al., 2019; Clabaugh & Matarić, 2018; Tapus & Mataric, 2008). As a particular type of assistive robotics, socially assistive robotics aims at helping users. However, given that socially assistive robotics is also a particular type of social interactive robotics, this help is provided resorting specifically to social interactions. Despite having different applications, both socially interactive robotics and socially assistive robotics stress the importance of generating realistic, effective, and human-like social interactions (Feil-Seifer & Matarić, 2005). In summary, socially assistive robotics shares assistive robotics' motivations but follows the same approaches as socially interactive robotics.

Some examples of fields where SARs are found to be especially effective include health care (such as therapy and rehabilitation, Tapus and Mataric, 2008), education (Clabaugh et al., 2019), daily life support (Kidd & Breazeal, 2008; Rodríguez-Lera et al., 2018) and therapy for children with cognitive disorders, such as Autism Spectrum Disorder (ASD) (Scassellati et al., 2018). In these fields, SARs can help not only by decreasing the workload of caregivers, tutors, and families (Clabaugh & Matarić, 2018), but also by engaging in personalised tasks that are beneficial to the users and would be more difficultly carried out by a human assistant (Clabaugh et al., 2019; Scassellati et al., 2018). For example, a SAR can generate a wide range of exercises disguised as games, choreographies or stories in order to galvanise children into engaging in such activities (Feil-Seifer & Matarić, 2005).

Feil-Seifer and Matarić (2005) stress that the effectiveness of SARs should be measured based on the performance of the user in accomplishing the tasks or activities in which the robot must assist the human. Additionally, SARs should be completely autonomous, should adapt to changes that occur in the environment and behaviour or mental state of the user, and should personalise their control system according to different characteristics, needs, and routines of the users. Finally, although there may be no need to physically assist the user, the robot's physical presence is still important to simulate social interactions that generate a social bond between the robot and the user (Tapus & Mataric, 2008).

Nonetheless, building and placing fully autonomous SARs in real-life environments where these robots are meant to assist users often brings social, scientific, engineering and logistic challenges. For this reason, a large part of SAR studies are conducted in laboratories rather than in the real world (Kidd & Breazeal, 2008). This is the case of the study carried outby Rodríguez-Lera et al. (2018), later described in Section 4.2. These challenges, which are described in the present chapter, require novel solutions. Therefore, a different approach from the ones found in the literature is proposed at the end of this chapter.

All in all, SARs resort to social interactions to improve the quality of life of users, by providing emotional and mental aid and support for people, and by taking over or reinforcing part of the work of doctors, therapists, nurses, family members or educators that need high investments of their time and energy (Clabaugh & Matarić, 2018).

4.2. Nature of Interactions

As mentioned in Section 4.1, SARs can provide aid to users in various areas, namely health care, education, specialised assistance to social groups with special needs, and daily life assistance. Since SARs can be used for many purposes, it is important to clearly define their role in each application, in order to select suitable characteristics and features - both physical and software-related - for the SAR (Feil-Seifer & Matarić, 2005). For example, the appearance and personality traits chosen for the robot must suit the complexity of the abilities and skills that the robot can display, in order to ensure that the users have sensible expectations from the SAR (Leite et al., 2013; Tapus & Mataric, 2008).

Based on the different goals and tasks that SARs are expected to help users with, several types of interactions between humans and SARs have been identified. In this section, in addition to covering the most common areas of applications of SARs, different researches on SARs including their aims, the roles taken by SARs, their physical characteristics and the type of tasks they assist their users with are analysed.

Regarding health care, socially assistive robotics in literature mainly focuses on rehabilitation and therapy. For example, several studies employed SARs in post-stroke rehabilitation including (Tapus & Mataric, 2008), which we briefly discuss next.

Tapus and Mataric (2008) focused on helping post-stroke patients by designing a SAR-coach that is capable of displaying a personality (more specifically, regarding introversion or extroversion and nurturing or challenging), showing empathy through verbal and non-verbal communication, and interpreting physiological cues including galvanic skin response and body temperature. The methods used to implement these three characteristics are described in Section 4.3. By adapting its therapeutic style to the user's personality and preferences, the SARs aimed to adapt to the fluctuations in the users' behaviours that occur from session to session, while being able to engage the user in long term. The tasks required from the patients consisted mainly in rather simple physical exercises that are deemed to be typical in the majority of post-stroke therapeutic interventions, including drawing direct lines on an easel, lifting books from a desk to a shelf, moving pencils between two bins, and turning pages of a newspaper. The exercises were coached by the SAR. To do so, the robot expressed both instructions and motivational statements to the user through vocal content. The style of these contents was mediated by the robot's personality, which was adapted to each user.

The interactions were divided into two phases, each one composed by one session. In the first phase, the duration of the session was between 5 and 8 minutes, while in the second phase the session lasted for 15 minutes. During the first phase, the user was guided by the robot to perform the aforementioned exercises. Prior to this phase, the users had filled in a questionnaire in which they evaluated their own personality, including their own level of introversion. Then, during the session of the first phase, the robot's level of introversion or extroversion was defined to match the user's. The effectiveness of this approach was assessed by qualitative answers provided by the users after the first session finished, i.e., the users recognised the robot's level of introversion to be similar to their own level. In the second phase, rather than identifying the user's personality traits, the parameters that defined the SAR's personality were tuned to maximise each user's performance in the task of transferring objects between two bins. By keeping track of how many times per minute the user could perform the transfer during the 15 minutes of this session, the robot was able to objectively track the patient's performance and adapt its behaviour accordingly. The techniques and algorithms employed by Tapus and Mataric (2008) to adapt the robot's personality (i.e. via adjusting the parameters of the SAR's personality model) to the user's performance are described in Section 4.3. Although the tasks required from the users in this research were physical, the attained results showed that the social and emotional aspects of post-stroke rehabilitation were crucial for an effective recovery. Nevertheless, despite stating that the robot should be able to engage the user's attention in a long-term scale, only short-term interactions were carried out when this robot was implemented to real-life experiments.

Although the majority of health care related SAR research focuses on therapy or rehabilitation, some health care researches have employed SARs to promote a healthier life-style without targeting a specific group. One of the first works to attempt this was carried out by Kidd and Breazeal (2008) with the goal of assisting users to lose weight. By creating a relationship between the user and the SAR, the researchers aimed at galvanising users to engage more often and effectively with their weight loss program. This research project had two main objectives: Firstly, to verify whether human-SAR interactions in long term can be successful and to identify the specific points that can be incorporated in human-SAR interactions to improve these interactions. Secondly, to investigate the usefulness of SARs in this particular health care domain. i.e., in weight loss programs.

The designed robot Kidd and Breazeal (2008) used had four degrees of freedom, a moving head with moving eyes, a camera, and a touch screen display. While the camera was used by the SAR to track the position of the user's face, the user could use the touch screen to provide information to the robot. The participants were divided in three groups with the same number of people, where the users from each group interacted with the weight-loss programme in a different development stage of the system, as to study the effects of the features added in between the stages. More specifically, the first group of participants (a control group) only used a paper log to maintain their data. Subsequently, a initial prototype was developed and one third of the participants interacted with the SAR for 6 weeks. Finally, the SAR was upgraded by adding the camera that allowed face-tracking, the moving eyes, and the capability to speak the text out loud, and the interactions with the last group were performed. All the participants interacted with the robot once or twice per day, throughout 6 weeks. Each session, which had an average length of 5 minutes, had the goal to assist the users to fill in the data regarding their progress. However, not only did the robot collect the users' data and guide them throughout the tasks, but also enriched the interaction through expressive gestures and small talk, giving advice and suggestions. The conversation script was generated based on several personal and temporal variables, which will be described in detail in Section 4.3. All this information contributed to personalise the interactions with respect to every user and to adapt the interactions according to the current needs of a user. As a result, each interaction between the SAR and a user could be unique. Moreover, the robot's speech was crafted in a way that it expressed positive and nurturing feelings as to motivate the user. The results showed that the participants that received the SAR used it more often and for a longer period of time than the participants who received traditional methods, such as digital or paper data-logs, used their programmes. Moreover, the robot was able to create a personal bond with the users and lead them to trust it. In addition, given that the duration of this study was significantly lower than the amount of time required for weight loss programs to achieve meaningful results, the the users did not lose significantly more weight using this approach when compared to other successful prior approaches. This results were obtained from the analysis of the data introduced by the user, the recordings of the dialogue between the robot and the user, and the user's opinion collected in a final interview.

In education, SARs have many benefits. For example, these robots can generate specialised teaching methods to respond to the unique reasoning and learning process of every child, yielding a more effective and efficient knowledge transfer procedure. Although it is not expected that SARs take the roles of the teachers all by themselves, they can act as auxiliary tutors, to provide individualised education and to allow each child to be more thoroughly accompanied (Feil-Seifer & Matarić, 2005). Furthermore, SARs help students not only to improve their cognitive abilities, but also to promote social interactions with children from different social groups. In the field of educations, SARs can act as educators, coaches, and peers (Clabaugh et al., 2019; Feil-Seifer & Matarić, 2005; Leite et al., 2013). As remarked by Leite et al. (2013), children are more easily distracted and may lose interest earlier in education-related activities than adults. Therefore, retaining the attention of the user throughout the interaction (commonly referred to as *user engagement*) is of added importance when dealing with children. Thus, SARs who assist children, including the robots designed for educational purposes, are specially focused on retaining the user's attention and use.

Clabaugh et al. (2019) studied the effects of long-term interactions between SARs and children with ASD, by considering to personalise the robot's behaviour to each user. Having the specific goal of improving the children's skills in mathematics, the robot proposed different maths-related games to the users. Resorting to the consultancy of pedagogic specialists, these games were designed not only to develop addition, counting, and pattern matching abilities of these children, but also to be suitable to the developmental needs of children with ASD within the considered age gap. Seventeen children of ages 3 to 8 years old participated in this project. The stations in which the children interacted with the SARs were installed in the house of every user and included a touch screen, a computer, a camera, and the "SPRITE robot", shown in Figure 4.1. This robot was fixed to the table and could move with six degrees of freedom in order to display "child-like body movements" in response to the child's success or mistakes when playing the games. Moreover, in order to show different facial expressions, the face of the robot was illustrated by a screen as a set of two eyes, two eyebrows, and one mouth (see Figure 4.1).

Instead of playing the role of a tutor, the SAR was designed to interact with the children as an assistive friend who provided guidance on the proposed games. Thus, the SAR introduced itself to the user as Kiwi, a space explorer that was lost and needed to go back to its planet. Kiwi told the children that they could help it to find its way back by playing the games that were being presented on the touch screen. The amount of feedback given by the SAR to a user and the difficulty of the games that were offered to every user were adapted to the performance of every child, in order to retain their engagement and to effectively improve their mathematical skills. The control approaches and algorithms used to implement the adaption process of Kiwi are explained in Section 4.3.

The interaction of Kiwi with every child had a minimum length of 30 days including 20 sessions. The children were suggested to engage in 5 interactions per week, performing 10 games during each interaction. The SAR was able to autonomously interact with children in their own house for an average of 42 days. The user's engagement was determined by analysing the video data of the interactions to assess the attention of the child to the SAR during the interactions. Furthermore, the evolution of the mathematical skills of the children was measured based on the scores they attained while playing the

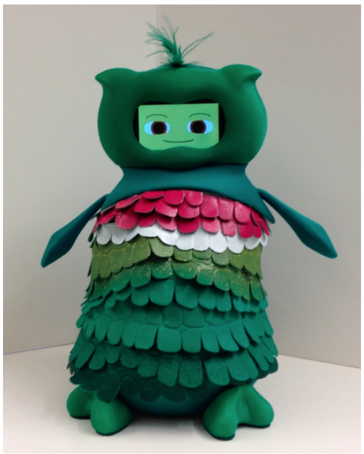


Figure 4.1: Sprite robot used by Clabaugh et al. (2019) to study the effect of long-term interactions between SARs and children with ASD. Source: Jordan et al. (2019)

games, by comparing the scores obtained before and after the entire long-term interaction. Moreover, the usefulness, effectiveness and adaptiveness of the SAR were assessed according to the results of the questionnaires that were filled in by the families. Contrarily, the relationship of the SAR with the children was assessed based on the surveys answered by the users. The results showed that, in average, the attention of the children to the SAR throughout the entire long-term interaction, as well as throughout every session, remained constant. Moreover, asides from the children for whom the games were too difficult or too easy (i.e., the youngest and oldest children, respectively), the mathematical skills of the users improved. The users who found the SAR to be more adaptive also found it to be more useful, according to the questionnaires filled by the families.

Nonetheless, the support provided by SARs for children is not limited to education. One the one hand, SARs have shown to be successful and effective in accompanying and aiding children with ASD to develop or improving skills that these children usually lack, including communication and social skills. On the other hand, SARs are able to help care-givers and families by reducing their workload. SARs can directly interact with children, propose personalised tasks to them and, sometimes, reach better results in improving their social and verbal skills than the ones attained by human helpers (Scassellati et al., 2018). In general, the use of SARs is not only well accepted by children with ASD, but also results in helping them effectively to overtake some educational and social barriers that are normally provoked by their disorder (Clabaugh et al., 2019; Scassellati et al., 2018).

Similarly to Clabaugh et al. (2019), Scassellati et al. (2018) focused on the interaction of SARs with children with ASD in an at home setup, where the robot was fully autonomous. The 12 children who participated in the study interacted with the robot for 30 days, 30 minutes each day. Contrarily to Clabaugh et al. (2019), who focused on aiding the children with educational skills, Scassellati et al. (2018) proposed a personalised robot that aimed at improving the emotional and social skills of the users. The motivation behind the goal of this research is the deficit in various social and communication skills that is a characteristic of children with ASD. The social skills considered in this study comprised joint attention, "social and emotional understanding, perspective-taking, and ordering and sequencing" (Scassellati et al., 2018). Joint attention is a non-verbal social skill where one rational agent calls the other agent's attention to an object by eye-gazing or pointing at it. Moreover, this skill is generally underdeveloped in children with ASD and important in social interactions. The interactions consisted of a story-telling moment guided by the SAR, followed by three games. The story aimed at developing social and emotional understanding skills by challenging the children to describe or guess the emotions or beliefs held by the characters at certain points in the narrative. The games were carefully crafted into exploring the aforementioned perspective-taking, ordering and sequencing skills, while teaching other concepts to the children, such as turn-taking.

The at home setup included a SAR (see Figure 4.2), a touchscreen monitor that the children could use to interact with the robot, and two cameras. Moreover, a third person, the caregiver, was present during the interactions. One camera tracked the attention foci of both the child and the caregiver by

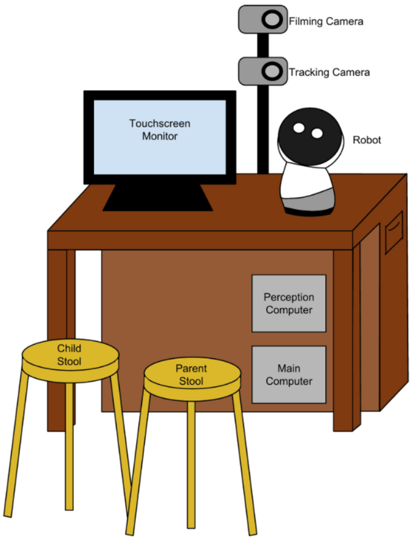


Figure 4.2: Setup used by Scassellati et al. (2018) in a long-term, at home interaction between the SAR that is displayed in the figure and children with ASD. A monitor was located next to the SAR, which allowed the child to interact with the robot. Two stools were placed in front of the monitor and the SAR, to allow the child and the caregiver to seat down while interacting with the robot. Additionally, two cameras were present: while one recorded the interaction for posterior evaluation, the other tracked the attention foci of the children and the caregiver. Source: Scassellati et al. (2018)

analysing the direction of their heads. In this context, the attention foci of the caregiver was necessary to assess the joint attention skills of the child. The other camera recorded the entire session for posterior evaluation. The robot expressed its behaviours through body motion, colour-changing lights, and eyes which could both blink and dilate. By maintaining eye contact with the child throughout the interaction and occasionally gazing at the screen, the robot exemplified engagement and joint attention, respectively. The progress of the children, which was used to indirectly measure the effectiveness of the SAR, was assessed based on, firstly, the children's gaze focus (to evaluate their joint attention skills), reaction to their name being called, or a greeting directed to them, and secondly using a questionnaire that was filled in by the caregiver who had participated in the session. The results demonstrated an improvement in the children's joint attention and communication skills. Furthermore, they proved the ability of the SAR to engage the children almost constantly during the entire duration of the study, i.e. throughout the 30 days (Scassellati et al., 2018), which is usually believed to be an issue in long-term interactions with children (Leite et al., 2013), as discussed in Subsection 4.4.2.

Finally, SARs can act as companions and assistants in the user's daily life, without targeting any group of users in particular. One example includes the work by Rodríguez-Lera et al. (2018). Given that the actions illustrated by SARs in the literature after a number of interactions become repetitive, Rodríguez-Lera et al. (2018) proposed a control architecture for a SAR that generated more spontaneous and natural behaviours. The SAR's role was to act as a companion and assistant for a regular person in their daily-life. The SAR could take on diverse roles related to home assistance and companionship, while displaying different decisions and actions depending on the selected role. In a first phase, the tests were conducted in a simulated setup. Subsequently, the developed controller was implemented on a real robot which interacted with humans in a laboratory environment. In both phases, the following scenario as considered: the SAR ran a quiz composed of 20 questions with the human (the SAR was responsible to store the relevant data and to lead the quiz). While the SAR and the human were interacting for the quiz, the doorbell rang, which prompted the robot to attend it. These tasks allowed to assess the verbal communication of the SAR, as well as navigation and context recognition. The experiments showed that the main goal of the project was achieved, i.e., by engaging in different roles, the SAR behaved slightly different in the presence of the same external inputs or, in other words, the SAR displayed a non-repetitive behaviour.

This section presented an overview of the nature of the interactions in which SARs and humans commonly engage. Various researches that were carried out in contrasting settings and for dissimilar time amounts were described: while in some of these researches the SARs accompanied the users in their houses for longer terms, i.e., more than a month, other experiments were carried out only in laboratory environments and their interactions lasted for shorter terms, i.e., a few minutes. In these researches, SARs were considered to assist different people to improve their physical health in rehabilitation programs, their skills in particular tasks (e.g., mathematics) in education programs, or their life style and well-being in general. Furthermore, the robots designed or employed as SARs in these reserches took various roles, including coach (Kidd & Breazeal, 2008; Tapus & Mataric, 2008), peer (Clabaugh et al., 2019), assistant and companion (Rodríguez-Lera et al., 2018). Finally, the tasks that were proposed by SARs to the users were either physical or cognitive. The physical exercises were proposed by Kidd and Breazeal (2008) and Tapus and Mataric (2008), within the scope of health care and rehabilitation. The remaining tasks were either cognitive, such as mathematical exercises (Clabaugh et al., 2019), or social, such as exercises to develop communication and emotional skills (Scassellati et al., 2018). Although the suggested tasks ranged from physical to cognitive or emotional, the human-SAR interactions in all these researches were social interactions rather than just physical ones.

The researches that considered the user of SARs in health-related applications displayed the positive impacts that SARs can have both for users in therapeutic or rehabilitation applications and for users who do not require any kind of medical intervention but still want to improve their life style and well-being through the assistance of SARs. In both cases, SARs have shown to successfully motivate people to work towards attaining their goals. The researches that focused on education and assistance for children with ASD stressed the importance of personalisation, user engagement, and long-term interactions in order to achieve good results via SARs. As long as these aspects are well tackled, SARs are capable of having significant impacts on the lifestyle and learning procedure of children with ASD.

4.3. Control Approaches for Steering SARs: Theory & State-of-the-Art Examples

This section describes the main approaches and control techniques for SARs that generate the state-of-the-art human-SAR interactions. The majority of the research projects in the literature resorted to rule-based (Kidd & Breazeal, 2008; Scassellati et al., 2018) and Artificial Intelligence (AI) approaches, such as Reinforcement Learning (RL) (Clabaugh et al., 2019; Tapus & Mataric, 2008) and fuzzy logic control (Kahraman et al., 2020). Since the typical fields where current AI approaches thrive have a lower complexity and dimensionality than the field of human social interactions, the research projects that apply such techniques in this domain require human input or feedback to be successful (Clabaugh & Matarić, 2018). This section is divided into two parts, where the first part focuses on the theoretical principles behind the approaches employed in the literature, and the second part analyses how several research projects have employed these control approaches to steer the SARs.

4.3.1. Theory Background

The theory behind the control approaches used to steer the state-of-the-art SARs is explained in this section. Initially, a theoretical explanation of Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs) is given, since these frameworks set the basis for Reinforcement Learning (RL), an approach often used in socially assistive robotics. Subsequently, the fundamental principles of RL are explained, in particular two RL algorithms: Q-learning and Policy Gradient RL. Finally, the basic characteristics of rule-based systems and the theoretical principles behind Fuzzy Logic Control (FLC), which is a more specific rule-based control approach, are described.

Markov Decision Processes (MDPs). MDPs describe the interactions between an agent and the environment in which the agent resides. The agent can access the environment's state, which also comprises information regarding the agent's state with respect to the environment, through observations. Subsequently, based on the observed state, the agent takes an action that alters the state of the environment. However, the effect of the actions taken by the agent may not be deterministic (Kaelbling et al., 1998).

Furthermore, Markov processes, including MDPs, fulfil the *Markov Property*. This property states that the values of the state variables at time step $t + 1$ only depend on the values of the same state variables at time step t. In other words, the future of the system only depends on the present, being independent from the past (Kaelbling et al., 1998).

An MDP is defined by a tuple $< \mathcal{S}, \mathcal{A}, T, R >$, where:

- $\mathcal{S}$ is the state space set.
- $\mathcal{A}$ is the action space set.
- $T(s, a, s')$ is the state transition function, which gives the probability of transitioning from state $s \in \mathcal{S}$ to state $s' \in \mathcal{S}$ when taking action $a \in \mathcal{A}$.
- $R(s, a)$ is the reward function, which yields a value that represents the benefit of taking action $a \in \mathcal{A}$ in state $s \in \mathcal{S}$.

Due to the Markov property, at any time step, the next state and the obtained reward depend solely on the current state and on the action that is taken at that same time step (Kaelbling et al., 1998). Given this structure, the agent tries to maximise the reward obtained. Nonetheless, most of the initial states do not allow the agent to achieve its goals by executing only one action. Instead, the agent needs to execute sequences of actions to fulfil its desires, even if success is only attained after several steps. Nonetheless, the agent still wants to achieve its goals as soon as possible. Thus, the agent tries to select the actions that maximise not the immediate reward, but the expected weighted sum of the reward corresponding to the present and future time steps, in which the most immediate rewards have a higher weight than the future ones. This weighted sum is called *discounted return* and is displayed in the following equation (Kaelbling et al., 1998):

$$G_t = \sum_{0}^{\infty} \gamma^t r_t, \qquad 0 < \gamma < 1 \tag{4.1}$$

where G_t is the discounted return for time step t and r_t is the reward received at that same time step. The discount factor (γ) regulates the discounted return, defining how much future rewards influence this sum compared to the most immediate ones. The larger this factor, the larger the influence of the estimated future rewards on the discounted return ((4.1)).

In order to maximise G_t, the agent can learn a certain policy $\Pi : \mathcal{S} \rightarrow \mathcal{A}$ that defines the action to be taken in any state of the state space (Sutton & Barto, 2018). The expected discounted return obtained when the agent starts in state s and executes policy π is called the *value function* and is defined by (4.2). The value function measures how well a certain policy is capable of maximising the discounted return:

$$V_\pi(s) = R(s, \pi(s)) + \gamma \sum_{s' \in \mathcal{S}} T(s, \pi(s), s') \cdot V_\pi(s') \tag{4.2}$$

Therefore, the agent's goal is to compute the policy that, for each state, yields the action a that maximises the value function. This policy is called the *optimal policy* π^* and is given by the subsequent equation:

$$\pi^*(s) = \underset{a \in \mathcal{A}}{\text{argmax}} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} T(s, a, s') \cdot V_\pi^*(s') \right] \tag{4.3}$$

Finally, the value function obtained by employing the optimal policy, which is called the *optimal value function*, is defined as follows:

$$V_\pi^*(s) = \max_{a \in \mathcal{A}} \left[R(s, a) + \gamma \sum_{s' \in \mathcal{S}} T(s, a, s') \cdot V_\pi^*(s') \right] \tag{4.4}$$

MDPs assume that agents can fully observe the entire environment at any given time. Nonetheless, in many systems, this assumption prevents the system from being accurately described. POMDPs aim at addressing this pitfall.

Partially Observable Markov Decision Processes (POMDPs). POMDPs are a planing framework useful to model systems in which the agent cannot fully observe the state space and, for that reason, there is an inherent uncertainty regarding the current state space (Pineau & Thrun, 2001). The underlying concept of this framework is that the agent does not make decisions based on the actual state of its environment, but rather based on a distribution of the agent's belief about the state space that is constantly updated according to the agent's observations and actions. The belief state b of the agent is a discrete probability distribution over the state space $\mathcal{S}$ that represents for each state the agent's belief that the agent is currently occupying that state (Foka & Trahanias, 2007).

POMDPs are defined by a tuple $< \mathcal{S}, \mathcal{A}, T, R, \Omega, \mathcal{O} >$, where:

- $\mathcal{S}$, $\mathcal{A}$, T and R correspond to the state space, action space, state transition function, and reward function which altogether characterise an MDP;
- Ω is the observation set, which encompasses all the observations that the agent can make in the environment;
- $O(s', a, o)$ is the observation function, which yields the agent's probability of making each observation $o \in \mathcal{O}$ when action $a \in \mathcal{A}$ was taken and, as a result, the agent ended in state $s' \in \mathcal{S}$.

Starting from an initial belief state b_0, the update of b is based on the current/realised real state s, the action taken a, and the observation o by the agent. Thus, the belief state for the next state is given by:

$$b_{t+1}(s') = P(s'|o, a, b_t) = \frac{\mathcal{O}(s', a, o) \sum_{s \in \mathcal{S}} T(s, a, s') \cdot b_t(s)}{\sum_{s' \in \mathcal{S}} \mathcal{O}(o, s', a) T(s, a, s') b_t(s)} \tag{4.5}$$

Just like in an MDP, in a POMDP the agent maintains its goal to maximise the discounted return (Kaelbling et al., 1998). Nonetheless, since the agent can exclusively access the belief state b rather than the realised state s, the reward must be evaluated as a function of the belief state b instead of the real state s. For this reason, the discounted return, the value function, and the policies must be defined as functions of the belief state b and not of the state s (Foka & Trahanias, 2007). As opposed to (4.4), which defines the optimal value function in an MDP, (4.6a) defines the same function for a POMDP:

$$V_\pi^*(b) = \max_{a \in \mathcal{A}} \left[\rho(b, a) + \gamma \sum_{b' \in \mathcal{B}} \tau(b, a, b') \cdot V_\pi^*(b') \right] \tag{4.6a}$$

$$\tau(b, a, b') = p(b'|a, b) \tag{4.6b}$$

$$\rho = \sum_{s \in \mathcal{S}} b(s) R(s, a) \tag{4.6c}$$

with $\mathcal{B}$ being the set of all possible belief states (Foka & Trahanias, 2007). Similarly, (4.6a) resorts to a transition function and reward function defined based on the belief state rather than on the state. These two functions are defined as shown in (4.6b) and (4.6c), respectively.

Reinforcement Learning. Reinforcement Learning (RL) is a machine learning approach built upon the theory of MDPs (Barto & Mahadevan, 2003). In RL, the agent interacts with the environment in order to autonomously discover the optimal policy to achieve its goals (Sutton & Barto, 2018). MDPs define the basic rules and structure of the environment where the RL agent performs and with which it interacts (Barto & Mahadevan, 2003). However, since the definition of the optimal policy presented in the MDP framework depends on the transition function (see (4.3)), which is a stochastic function, the MDP theory does not present a way to compute the optimal policy in practice. Therefore, RL focuses on computing the optimal policy that an agent must follow to achieve its goals by means of allowing the agent to interact with the environment (Barto & Mahadevan, 2003). Similarly to human learning, the agent is expected to understand the consequence of its actions by trying different actions and receiving the rewards associated to them (Sutton & Barto, 2018). The more the agent interacts with the environment, the more it is expected to learn which are the most favourable actions in each circumstance.

Several RL algorithms have been developed throughout the years. Q-learning is one of the most used RL algorithms (Barto & Mahadevan, 2003) and is a baseline for more recent and complex RL algorithms. Therefore, it will be subsequently described.

Q-Learning. Similarly to many RL algorithms, Q-learning aims at maximising the *action-value function* $Q_\pi(s,a)$, rather than the value function $V_\pi(s)$ (Sutton & Barto, 2018). The action-value function, also named Q-function, yields the discounted return obtained when the agent takes action a in state s by following the policy π. The main difference between the action-value function $Q_\pi(s,a)$ and the value function $V_\pi(s)$ is that the first yields the discounted return for each pair of state and action (s,a) when the agent follows the policy π, while the value function yields the discounted return when the agent is in state s and follows the policy π. The action-value function is given by the following equation (Barto & Mahadevan, 2003):

$$Q_\pi(s,a) = R(s,a) + \gamma \sum_{s' \in \mathcal{S}} T(s,a,s') \cdot \max_{a' \in \mathcal{A}} Q_\pi(s',a') \tag{4.7}$$

The optimal action-value function $Q^*_\pi(s,a)$ defines the discounted return for each pair of state and action (s,a) when the agent follows the optimal policy π^*. In order to determine the optimal action-value function, Q-learning starts with an initial action-value function and, by integrating the information collected by the agent when interacting with the environment, this algorithm gradually updates the estimate of the action-value function (Barto & Mahadevan, 2003). By having the agent interact with the environment and update the current estimate of the action-value function in each step, the estimate eventually converges to the optimal action-value function (Sutton & Barto, 2018).

In practice, the agent executes an action a when in state s. The choice of the action often follows an ϵ-greedy policy. According to the ϵ-greedy policy, the agent chooses the greedy action (i.e., the action that yields the highest value of the currently estimated action-value function for that state) with probability $1-\epsilon$, or a random action with probability ϵ, where $\epsilon \in [0,1]$. As a consequence of executing action a, it receives an immediate reward r and observes the subsequent state s'. The estimate of the action-value function, $Q_k(s,a)$ is then updated based on the obtained immediate reward r and the observed next state s', according to the following equation (Barto & Mahadevan, 2003):

$$Q_{k+1}(s,a) = (1-\alpha_k)Q_k(s,a) + \alpha_k \left[r + \gamma \cdot \max_{a' \in \mathcal{A}} Q_k(s',a') \right] \tag{4.8}$$

where α_k is the learning rate parameter. This parameter controls how much of the old estimate is kept and how much is replaced with the new estimate. Generally, the learning rate is set to decay as the number of iterations increases (Barto & Mahadevan, 2003). It is important to notice that the update of the estimate of the Q-function does not depend directly on the policy. Nevertheless, the policy influences the choice of the action taken and, consequently, which pair of state and action (s,a) is updated in the estimate of the Q-function in each iteration (Sutton & Barto, 2018).

Q-learning is a simple algorithm suitable to deal with small and discrete action and state spaces. Recently, due to the need to deal with complex systems that possess high dimensional and/or continuous state and action spaces, new RL algorithms have been developed. One of these algorithms is Policy Gradient RL, which will be subsequently explained.

Policy Gradient RL. Policy Gradient algorithms apply a different approach than Q-learning to achieve the optimal policy. Rather than trying to estimate the action-value function and updating that estimate in each iteration, they describe the policy by several parameters, and perform a search in the space of the policies defined by those parameters. In each iteration, instead of directly choosing the action that returns the highest value of the action-value function, Policy Gradient RL estimates in which direction the parameters that describe the policy must be updated in order to improve the policy (Sutton & Barto, 2018).

Given the vector θ that comprises all the parameters that describe the policy, the gradient of a function ρ that reflects the performance of the current policy with respect to the vector of the parameters is computed. This function ρ is usually defined based on the reward function. Subsequently, the vector with the parameters is updated proportionally to the gradient, according to the following equation (Sutton et al., 2000):

$$\theta_{k+1} = \theta_k + \alpha_{PG} \frac{\partial \rho(\theta)}{\partial \theta}\Big|_{\theta_k} \tag{4.9}$$

where α_{PG} is a positive value that establishes how much the policy is updated in each iteration. By repeating this process, the vector of the parameters is expected to converge to a local maximum of the function that reflects the policy performance. The optimal policy is given by the parameters of the vector θ in this local maximum (Sutton et al., 2000).

Rule-based Control. In some systems, human experts have the knowledge of which action is the most suitable to be applied in each circumstance. Rule-based control aims at defining sets of rules based on expert knowledge, as well as the situations in which each rule must be applied, in order to control systems (Hayes-Roth, 1985).

Rule-based control express the rules in terms of *if-then* statements, with the following structure:

If *antecedent*, then *consequent*.

When one rule is triggered, its antecedent is assessed. If the antecedent can be satisfied, then the consequent is executed (Hayes-Roth, 1985).

A basic rule-based system consists of two parts: a part that stores information and another part that processes it. The storage block includes a database that stores the temporary data, and the knowledge base. The knowledge base comprises the set of rules used to control the system, as well as the facts, i.e., true propositions that are permanent. Both the temporary data and the long-term facts can be used to assess the antecedent of rules. The block that processes the information, which is called the *inference engine*, alternates between selecting rules and executing the selected rule. During the execution process, the antecedent of the rule that was selected is interpreted considering the data available in the temporary database and the facts from the knowledge base (Hayes-Roth, 1985).

Depending on the data introduced and/or available in the two databases, the system determines the most suitable sequence of rules. The rules are able to divide the state space in different sub-sections. Therefore, rule-based controllers are often able to deal with complex environments, as long as the rules that are necessary to describe the actions that must be taken in the different states of the system are included *a priori*. Moreover, rule based controllers allow transparent access to the logic behind their decisions, since all the decision process is described by the rules that are activated (Hayes-Roth, 1985).

Nevertheless, rule-based controllers become unpractical for very large and complex systems, given that it is too difficult to capture all the rules that are necessary to control the system, and predict all the circumstances that can occur. Moreover, in classic rule-based systems, the antecedent of the rules is assessed in terms of being totally true or totally false. However, in some systems, the premises should be assessed as a matter of degree. Therefore, classic two-valued logical systems cannot be used to control such systems accurately. Fuzzy logic control is a rule-based approach that is able to cope with both these pitfalls of classic rule-based systems (Zadeh, 1988).

Fuzzy Logic Control. Fuzzy logic, the basis of FLC, defines rules in a way that resembles the reasoning performed by humans (C. C. Lee, 1990; Zadeh, 1988). Fuzzy logic perceives and expresses the inexact, approximate nature of environments characterised by uncertainty (such as the real world) (C. C. Lee, 1990) and, consequently, is capable of taking rational decisions in such environments (Zadeh, 1988).

When fuzzy logic is used to model a rule base, the degree of fulfilment of the premises of the rules may vary within the range [0,1] compared to crisp logic, which implies that the premises of the rules are either totally true or totally false (Zadeh, 1988). Moreover, in FLC, the rules that control the systems are defined in linguistic terms. FLC is able to capture a linguistic control strategy that was developed based on expert knowledge, and adapt it to build an automatic controller that controls the system based on that knowledge (C. C. Lee, 1990).

One of the underlying concepts of FLC is fuzzy sets. Set U is a set of n objects u_i and is called *universe of discourse*. The fuzzy set A is defined in the universe U and is characterised by a membership function μ_A which can take any value in the range $[0, 1]$. Generically, a fuzzy set A is described by the following equation:

$$A = \sum_{i=1}^{n} \mu_A(u_i)/u_i \tag{4.10}$$

The other essential concept of FLC to define the rules is linguistic variables. Linguistic variables are defined by five elements: the name of the variable, denoted as x; the set of linguistic terms that the variable x can assume, named $T(x)$; the universe U in which the variable is defined; a syntactic rule that describes how the names of the values are created; and a semantic rule that associates the meaning of each element in $T(x)$ with the corresponding value of the variable x.

For example, considering the linguistic variable $x =$ temperature, the set of linguistic terms can be $T(\text{temperature}) = \{$“very low”, “low”, “mild”, “high”$\}$. In this example, a *very low* temperature can be considered a temperature below about $273\,\mathrm{K}$, while a *mild* temperature can be defined as a temperature close to $300\,\mathrm{K}$.

Finally, fuzzy control rules share the same structure as the conditions of the general rule-based system. Nonetheless, the antecedents and the consequents of fuzzy control rules are defined based on linguistic variables. As previously mentioned, the antecedents and consequents are assessed in terms of degree of fulfilment, rather than in terms of being completely true or completely false.

To build a fuzzy logic controller in order to steer a certain system, it is necessary to define the fuzzy sets, linguistic variables and fuzzy rules that are suitable for that system.

This section covered the theoretical background on the control approaches used to steer the state-of-the-art SARs. The applications of these approaches performed in the socially assistive robotics projects that were analysed in Section 4.2 are discussed in Subsection 4.3.2.

4.3.2. State-of-the-Art Examples

This section analyses the control approaches that were used in the SAR research projects described in Section 4.2, as well as how these approaches were employed to create controllers that aimed to achieve the goal of each research project.

In the scope of post-stroke rehabilitation, Tapus and Mataric (2008) used Policy Gradient RL to personalise the SAR to users. The authors started by researching the roles of both personality and empathy of the SAR in human-SAR interactions. Firstly, given the importance of the personality of humans in social interactions, Tapus and Mataric (2008) reasoned that assigning a personality to the SAR can be beneficial in human-SAR interactions. Thus, a parameterised personality that can be tuned according to both the user’s personality and the performance of the user was proposed. Secondly, based on previous researches, empathy has proven to positively influence the therapeutic process and “pro-social behaviours” (Tapus & Mataric, 2008). For a rational agent, having empathy for a human means understanding the state-of-mind, including emotions of the human. Empathy causes certain responses and reactions in humans, such as mirroring the body movements of the other human. Since the aforementioned reactions of empathy trigger positive social feelings, such as cooperation, Tapus and Mataric (2008) believed it to be possible to display empathy, not by simulating the process itself but its outcomes, i.e., by making the robot display the responses triggered by having empathy (such as mirroring the body movements of the human) rather than by allowing the robot to understand the user’s mental states. Finally, the interpretation of physiological cues, such as galvanic skin response and body temperature, is of utmost importance in the stroke-therapy context, since it contributes to a better awareness of the user’s inner state. The interpretation of this type of data allows SARs to adapt their behaviour to the user’s needs. More specifically, the physiological cues enable the robot to respond to short-term changes in the user’s inner state and to produce suitable empathetic responses to humans.

In Tapus and Mataric (2008), the personality and empathy of the SAR were simulated in the SAR’s behaviour through the choice of verbal expressions, proxemics (distance between the SAR and the user), the characteristics of the tasks proposed to the user, and other non-verbal cues. The non-verbal communication was controlled by the speed and complexity of the robot’s movements. Consequently, the robot was able to display different behaviours, simulating a personality that varied from extroverted/challenging to introverted/nurturing. Note that the parameters that characterised the SAR’s personality adopted values in a continuous range, allowing for a range or spectrum of personalities to be simulated for the SAR. As explained in Section 4.2, the SAR’s personality was tailored to each user in order to help them attain their best possible performance. Before engaging in the therapeutic exercises, the participants filled in two questionnaires. The first aimed at collect personal data, such as gender and age, and the second determined their personality traits, i.e. their positions on the spectra of introvert/extrovert and nurturing/challenging. Based on the premise that humans have more affinity to humans or rational agents that display a similar personality, the user’s personality was used as a starting-point to generate the SAR’s personality in the first phase of the interactions. The patients were also exposed to a random robot personality in order to allow a comparison between the effect of both personalities and extract conclusions. The results showed that the users deemed the first personality (which was selected to be similar to the user's) more similar to their own than the second personality (which was randomly selected). This result implies that the cues used to simulate it were effective. Moreover, users achieved a better performance when they were coached by a SAR with a similar personality.

Subsequently, in the second phase the robot's personality was more specifically personalised to the user considering the user's preferences, and using Policy Gradient RL. Starting from the initial policy from the first phase, the learning approach consisted in estimating the gradient of the reward function for the parameters that described the SAR's personality; subsequently, the computed gradient was employed to reach a local maximum, yielding a new policy which corresponded to a modified personality. This method was used for the entire session in the second phase (15 minutes), where the effectiveness of each policy was assessed for an interval of $60\,\mathrm{s}$. The reward function was defined based on the performance of the user, i.e., whether or not the goal of a particular task was achieved. The results substantiated the SAR's ability to adapt its personality and empathetic capacity to each user, since the performance of the users reached their peak when the SAR personalised its actions and personality. Tapus and Mataric (2008) stated that longer sessions than the ones considered might be required to reach optimal policies. Moreover, evaluating each personality for only $60\,\mathrm{s}$ is not long enough to gather representative results, since in larger time intervals fluctuations in the performance can occur due to other factors such as the user's fatigue. These critics reinforce the need for long-term interactions in order to achieve accurate personalisation. Furthermore, the underlying principle of empathy includes understanding the other agent's inner mental state, perceptions and reasoning process. Although it is possible to show empathy by displaying the reactions that this feeling triggers in humans without understanding the inner state of the user (as proposed by Tapus and Mataric (2008)), this approach cannot recognise the correct moment to show empathy to the user. Since the empathy process that leads to the displayed outcomes is not actually modelled, it is unlikely that this approach is able to accurately yield the correct outcomes in all circumstances. Consequently, for a rational agent to feel and accurately display empathy, it should take the perspective of the other individual into account and understand how external factors can be internalised by and influence the other individual. For this reason, modelling the agent cognition of the human with whom the SAR is interacting is expected to be highly beneficial.

The SAR that helped motivating users to lose weight proposed by Kidd and Breazeal (2008) (see Section 4.2) used a rule-based approach to generate a dialogue script based on variables such as the current time of the day, the stage and the evolution of the relationship between the SAR and the user, and the information that was provided to the SAR by the user via the touch screen. While some variables or user behaviours could trigger a certain type of speech in the SAR, other variables controlled the content of that speech. For example, turning on the device for the first time in a day lead the robot to greet the user (type of speech), while the time of the day defined the content of the greeting. Moreover, in order to estimate the relationship status, the following method was used. The relationship status could take one of the three following values: initial, normal, and repair. The *initial* status was used in the first days, in order to get the user acquainted with the system. In this status, the SAR presented thorough information to the user and adopted a more polite tone. The *normal* status was the default mode used after the initial period and, consequently, the robot employed a more succinct and direct language. The *repair* status was activated when the relationship was considered damaged. In this mode, the robot attempted to smoothly address the fact that the relationship was falling short of its potential, encouraging the user to be more active. In the first 4 days of the experiment, the relationship was set to *initial*; in the 4 following days, the relationship was set to *normal*. From the ninth day on (and until the end of the experiment), the SAR assessed the status of the relationship according to the answers that were provided by the user to a set of 8 questions. The questions reflected the trust and faith that the user had grown in its relationship with the SAR and the answers were given on scale from 0 to 600 [1]. Every day, two questions were selected from the set of the 8 questions and were asked to the user in such a way that, every four days, all the questions were asked exactly once by the SAR to the user. Every day, a formula was computed based on the answers that were provided by the user in the last 8 days. The scores given by the users in the previous four days were averaged, as well as the scores from the four days before that, yielding two averages. The oldest average was

[1] In this scale, 0 corresponded to the least and 600 to the most amount of trust and faith.

subtracted from the most recent one, yielding a final “relationship score”. If this final score was lower than a certain threshold defined *a priori*, the conclusion was that the relationship between the SAR and the user had progressed negatively, and the relationship status was set to *repair* by the SAR; otherwise, if the relationship score was higher than the specified threshold, the relationship status was set to *normal*. Subsequently, the SAR adjusted its behaviour based on the active relationship status. Both the generation of the dialogue text and the computation of the relationship status were rule-based methods, given that the text of the dialogue was generated based on scripts that were stored in the databases of the SAR, and the relationship status was set as a function of the described rules (Kidd & Breazeal, 2008). It is important to note that although the SAR assessed its relationship with the users via a mathematical model, the robot itself was controlled via a rule-based approach without actually modelling the user’s cognition and state-of-mind.

The study carried out by Clabaugh et al. (2019) employed RL to adapt the mathematical games presented by the SAR to each participant. Although this control technique required a large data-set when dealing with complex environments, the authors state that the long-term duration of the interactions considered in this research (at least 30 days) tackled this problem.

A hierarchical controller (see Figure 4.3) was implemented using a meta-controller in the highest level of control that chose one out of the five different actions, each one of them was regulated by a lower-level controller. The personalisation of the controller to every user was then carried out in two of the five lower-level controllers: the instructions controller and the feedback controller. In particular, the operation of the instruction controller was according to various Levels of Challenge (LoC), which refer to the degree of difficulty of the tasks that were proposed to children. The feedback controller also performed based on various Levels of Feedback (LoF), which refer to the amount of information or feedback that was provided for the children to assist them in accomplishing the tasks. Thus, the LoC and LoF were personalised to each user, based on task performance. Five different LoCs and four different LoFs were considered. The choice of the correct LoC and LoF was structured as an RL problem, which was trained using Q-learning. In both cases, the action space included all the LoCs and LoFs, and the state space consisted of the ten different games. In order to improve the mathematical skills of the children, the controller aimed to select the highest possible LoC for a child with ASD, while making sure that the child does not make too many mistakes. Accordingly, the reward function of the instruction controller was defined as a function of the proposed game, the value corresponding to the LoC, and the number of mistakes. The reward of the feedback controller was maximum for the lowest possible LoF that could still prevent that too many mistakes or help requests being made by the child. Accordingly, the reward of the feedback controller was a function of the proposed game, the value corresponding to the LoF, and the number of mistakes and help requests made by the child.

After a certain number of interactions and using Q-learning, both the LoC and LoF converged for the participants which reveals the importance of the long-term interactions to achieve effective personalisation. Finally, considering the two main objectives of the experiments, the engagement of the children remained constant throughout the entire study and their maths skills improved (Clabaugh et al., 2019).

Similarly to the research by Clabaugh et al. (2019), the work done by Scassellati et al. (2018) aimed to help children with ASD to develop skills through the employment of educational games. The personalisation was achieved with a rule-based approach: the level of difficulty of the games and stories proposed by the SAR to the children was adapted based on their performance in the previous game. If the children surpassed a score of 75%, the SAR increased the level of difficulty of the next games and stories, while if the score obtained by the children was lower than 25%, the SAR decreased the level of difficulty of the next games and stories. Finally, if the attained score was in between the two aforementioned percentages, the SAR maintained the same level of difficulty for the games and stories as before. The results of the experiments conducted by Scassellati et al. (2018) showed that the SAR was capable of maintaining the interactions and keeping the users engaged in long term (a one-month duration was considered overall in their experiments). Within the rule-based approaches, the aforementioned study carried out by Scassellati et al. (2018) stood out.

In a survey regarding the use of fuzzy logic in humanoid robots, Kahraman et al., 2020 noticed the potential benefits of applying this technique to model the behaviour of robots, remarking that little work has been done in this area. When using fuzzy logic control, the degree of fulfilment of the premises of the rules varies between 0 and 1. Moreover, the models that are based on fuzzy logic can include linguistic terms. These two characteristics of fuzzy logic are also present in the way humans reason

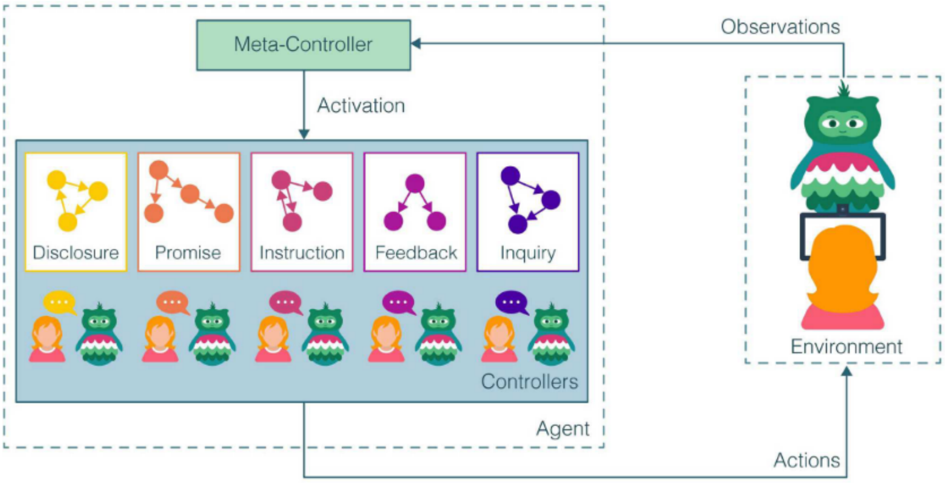


Figure 4.3: Hierarchical controller developed by Clabaugh et al. (2019). The meta-controller occupied the highest level in the hierarchy and had the objective of choosing one out of five actions. Each one of the five actions was controlled by a lower level controller. The personalisation to the users was implemented in the instruction and feedback lower-level controllers. Source: Clabaugh et al. (2019)

and plan. Therefore, although most of the SAR researches that employ fuzzy logic control use this approach for navigation purposes, Kahraman et al. (2020) propose that these rules are ideal to simulate human-like decision making processes in humanoid robots. Nonetheless, Kahraman et al. (2020) did not attempt to implement this suggestion.

The majority of the studies in literature use model-free approaches to control SARs. Nonetheless, Rodríguez-Lera et al. (2018) also discusses modelling of the cognitive processes behind human actions in order to generate more natural human-SAR interactions. For SARs to behave more human-like, Rodríguez-Lera et al. (2018) proposed to integrate the following two control approaches to develop the decision making module of SARs: (1) control based on a model that represents the procedures behind the actions of the SAR, and (2) control based on a mapping from the inputs to the SAR to its actions, without developing an explicit model for the SAR. The authors stated that the first approach alone lacks spontaneity in responding to unknown situations, while the second approach alone lacks the social knowledge behind the decisions that are made by the SAR. By integrating the two approaches, the two aforementioned flaws can be addressed. Moreover, in order to make the SAR's behaviour more believable and human-like, the SAR's actions and behaviours were designed to depend on its task, and on its personal goals and needs. For the needs, since they depend on the role the SAR takes, four roles were defined in this experiment for the SAR: semi-autonomous service, tele-operated service, companionship and autonomous service/assistance. These roles not only defined the set of needs but also the set of motivational variables, which reflect the SAR's will and drive to perform actions.

According to the priorities, a hierarchy of needs was defined for the SAR. The lower the level, the more basic and surviving-related the needs. For this reason, lower-level needs were more urgent and had priority when compared with higher-level needs. This project considered three levels of needs. The lowest, termed existence, dealt with basic existential necessities such as rest, fear, pain and safety. The middle one, named relatedness, comprised the SAR's social interactions and personal needs, such as comfort, curiosity, and frustration with the human. The growth needs (the highest level) encompassed the needs of learning and expanding the SAR's current knowledge. The SAR's behaviour was then produced based on a concept called pool of finite state machine (FSM). The pool of FSM included all sequences of behaviours of the SAR, represented by states and arcs. Each state represented the behaviour of the SAR at a particular time. From a state there were several transition arcs which lead to the next possible states. Each arc had three conditions dependent on the motivational variables, external stimuli, and behavioural scheduling. The arc whose conditions were true was activated, defining the next behavioural state. This process was repeated to generate sequences of actions and behaviours for the SAR. The behaviours were generated based on a tree-like structure, i.e., the high-level behaviours were subdivided in lower-level, simpler actions. These actions were then subdivided in even simpler commands that consisted in hardware inputs. The pool of FSM of the SAR was managed by three subsystems. The first, deliberative subsystem, managed long-term plans which required complex sequences of actions. For every goal, this subsystem generated a pool of FSM, i.e., the possible sets of course of actions according to the robot's role, motivational variables, needs, and inputs. The Reactive subsystem responded quickly to the environment. Despite its goal to generate spontaneous behaviours, the reactions proposed by this subsystem did not hinder the achievement of the long-term goals. The third was described by a set of variables, which depended on the SAR's role. Each motivational variable was described by its value, a hierarchy level, limit values, threshold values and a set of other associated variables. While the limit values defined the minimum and maximum values that the variable could assume, the threshold values managed the uncontrollable increase or decrease of the values of the variable.

Two experiments were conducted to evaluate this architecture, its implementation, and particularly the impact of the motivational subsystem. The first experiment consisted in computational simulations to tune the saturation values and limits of the motivational variables. The second experiment consisted of human-SAR interactions. All the experiments were conducted in a laboratory environment. The simulations were carried out following the same external conditions, varying the SAR's role only. The results proved that, even in the same circumstances, selecting a different role for the SAR induced small differences in the robot's behaviours. Furthermore, this system was capable of planning the robot's behaviour, executing the plan or adapting it if necessary, based both on external conditions and internal motivations of the SAR. Since the approach is model-based, there is access to the reasoning and motivations behind the robot's behaviours, which provides transparency for the proposed approach.

As previously mentioned, the majority of the studies comprised in this survey employed model-free techniques. Clabaugh et al. (2019) and Tapus and Mataric (2008) employed goal-based approaches, while Scassellati et al. (2018) and Kidd and Breazeal (2008) used rule-based approaches. Both the methods used by Clabaugh et al. (2019) and Scassellati et al. (2018) were dependant on the user's performance, and no user modelling was performed whatsoever. Moreover, the studies which comprised prior knowledge about the user in their model did it superficially. Tapus and Mataric (2008) based the SAR's actions on the user's level of introversion. Nonetheless, one personality trait is hardly representative of the user's uniqueness. In the work developed by Kidd and Breazeal (2008) the only aspect which was modelled was the status of the human-SAR relationship. However, the cognition processes behind the human's behaviour were not modelled. The aforementioned projects leaned on task-performance metrics in order to tune and evaluate the controllers that were used for SARs. In reality, only a few authors within the SAR research community resorted to user modelling. In particular, the research by Rodríguez-Lera et al. (2018) modelled parts of the human-like cognition procedure considering the needs and motivations. Nonetheless, this approach was applied to the SAR to develop a human-like personality for the SAR that influences the actions of the robot and makes them more human-like and transparent. To the best of our knowledge, however, a thorough model of the cognition of the human that interacts with a SAR, for the SAR to understand the human, and to adapt and personalise its actions accordingly has not been investigated in human-SAR interaction research.

4.4. Interaction Length

In the interactions between humans and SARs, it is important to make the distinction between short-term and long-term interactions. Although both types of studies share the fundamental goal of helping users overcoming or improving a physical or social aspect, the nature of the conclusions that can be taken from each type of interaction are usually different. While short-term interactions usually aim to test if a robot (or the introduction of a feature in an existent robot) can be helpful in a certain circumstance when compared to traditional methods (Rodríguez-Lera et al., 2018; Tapus & Mataric, 2008), long-term interactions focus on how an SAR can assist users over a longer period of time (Clabaugh et al., 2019; Kidd & Breazeal, 2008; Scassellati et al., 2018). Therefore, in the latter, SARs have to deal with more complex challenges, such as not being able to integrate themselves in the environments where they must aid the user, or losing the interest of the user (Leite et al., 2013).

Literature does not agree on the amount of time required for a study to be acknowledged as long-term. Considering the research projects that were analysed in the present literature survey, the shortest long-term interactions lasted for a month (Clabaugh et al., 2019; Scassellati et al., 2018).

Leite et al. (2013) stated that the number and duration of individual sessions is more relevant to define whether a study can be considered long-term than the duration of the study itself. Moreover, rather than establishing a value, the authors argue that the number and duration of the sessions comprised in each long-term study should depend on to the characteristics of said study. Examples of these characteristics include the number of users interacting with the robot at the same time (if more than one) or the amount and complexity of behaviours exploited to sustain users' engagement. A more clear definition of the recommended duration of a long-term project is given at the end of this section, since it is based on concepts discussed on Subsection 4.4.2.

This section comprises an analysis of short- and long-term interactions, the differences between the two, and their characteristics and pitfalls. Given the higher complexity inherent to long-term studies and the focus of the present survey on those, the studies which comprised long-term interactions are analysed more in detail: the most common challenges in long-term projects are identified and the approaches found in literature to tackle them are presented.

4.4.1. Short Term Interactions

The majority of the studies regarding SARs comprise merely short-term interactions. Whenever a new application for SARs or framework to control them is proposed, it is important to test its potential. Given the challenges inherent to employing SARs in long-term interactions and the resources required to overcome them, it is logical that many studies resort to short-term interactions as a first approach.

This was the case of the project carried out by Tapus and Mataric (2008), in which the longest interactions lasted for 15 minutes. The performed tests aimed at investigating the hypothesis that empathy, personality and physiological cues were important in human-SAR interactions and in the post-stroke rehabilitation domain. Nonetheless, the authors recognise that, in order to understand the full extent of the benefits brought by their robot and achieve an accurate personalisation, longer interactions must be carried.

Similarly, the tests performed by Rodríguez-Lera et al. (2018) only employed short interactions. Although their length was not quantified by the authors, the interactions followed the script described in Section 4.2, which was estimated to last approximately 2 minutes. These experiments had the objective to test whether the proposed control architecture was capable of generating different behaviours based on the robot's internal variables. They were carried out in laboratory and repeated under similar external conditions, only changing the values of said internal variables. Although the short-term tests achieved their intended purpose, the authors also stressed the necessity to test the proposed architecture in long-term interactions.

The reason why authors who used short-term interactions recommend repeating the experiments in longer-term settings is that short-term tests are not representative of the interactions that these robots should engaged with humans when fulfilling their role in the real-world. In most cases, if these robots truly aim at improving their users quality of life, they must interact with them on a constant basis. Moreover, since the robot's benefits tend to disappear as a result of extended exposure, the results obtained in short-term studies usually cannot be extrapolated to longer ones. For example, in some cases, the improvements that were observed during the experiments regarding the task performance of the users were found to be forgotten by the users once the interaction between the SAR and the user ceased (Scassellati et al., 2018).

For this reason, despite being good indicators of the potential of a certain technique or application, the results obtained in short-term interactions do not provide an accurate prediction of the human-robot interaction that can be expected once the SAR is assisting users in real-world environments (Scassellati et al., 2018).

4.4.2. Long Term Interactions

As mentioned, when compared to short-term interactions, long-term interactions are less common in the SAR domain. This is likely a consequence of the challenges brought by having a SAR behaving autonomously for a long period. Moreover the disparity in both time and resources required to carry out studies that comprise long-term interactions, when compared to studies that only include short-term interactions, is another potential cause for the shortage of studies that comprise long-term interactions (Leite et al., 2013).

The study of long-term interactions requires relevant tests that evaluate both the SAR and the evolution of the interactions between the human and the robot. Ideally, said tests must be performed in the real-world environment where the SAR is expected to help the user (Clabaugh & Matarić, 2018). Therefore, the study of long-term interactions between humans and SARs demands designing and building

SARs that are robust enough to reside in a real-world environment without human intervention, which inevitably brings scientific, engineering and computational challenges (Clabaugh et al., 2019).

Firstly, lengthy studies usually require a very prolonged pretesting phase in laboratory. These studies are important to ensure that the robot can behave correctly in an uncontrolled environment. Moreover, despite the complexity inherent to predicting in advance all possible human responses, the robot must be ready to adapt its behaviour to them. Thus, it is of utmost importance to either predict them or to provide the robot with a mechanism to deal with them. Additionally, the SAR must have enough autonomy to last for the entire duration of the study. Consequently, many long-term studies are carried out in laboratory instead of in the environments in where they are expected to assist their users (Leite et al., 2013).

Apart from the engineering challenges brought by long-term studies, the main barriers to attaining successful long-term interactions between humans and SARs are caused by social aspects (Clabaugh et al., 2019).

Some authors in the literature agree that ensuring the engagement of the users during the interaction is extremely challenging and that it was one of the main flaws in prior attempts to develop SARs capable of long-term interactions (Clabaugh et al., 2019; Kidd & Breazeal, 2008; Leite et al., 2013). As explained by Leite et al. (2013), the added difficulty in maintaining the attention of the users in long-term interactions, when compared to short-term interactions, is linked to the *novelty* effect. According to the authors, when users start to interact with a SAR, they are dazzled by its apparent sophistication and its features. Nonetheless, after some time, the user explores the unknown features and aspects that the SAR can offer. Subsequently, this novelty effect fades away and, if suitable techniques are not adopted, the robot loses the attention of the user.

User engagement and its evolution can be measured directly by the number of sessions in a certain time interval, how that frequency changes over time, and the time dedicated to the SAR in each session. It can also be measured by indirect cues, such as the amount of time spent by the user gazing at the robot (Leite et al., 2013).

Although different SARs have different purposes, they share the ultimate goal of positively impacting and improving the quality of life of their users. Therefore, despite the challenges and difficulties inherent to the development of SARs that are capable of long-term interactions, SARs must assist and accompany their users throughout the time that the users need, even if that time is indefinite (Matarić & Scassellati, 2016). In scenarios in which robots must assist users to overcome a deficit or issue, such as SARs that focus on therapy or rehabilitation, the interaction should last until the problem is solved. When the robots operate as a daily life assistant, they should be able to interact with the user indefinitely. In summary, the end of a human-SAR interaction should be defined by the needs of the user, rather than by the technological and social constraints of the SAR. For this reason, the research of long-term interactions, as well as the creation of new approaches and techniques that can tackle the challenges inherent to this type of interactions, is of utmost importance (Matarić & Scassellati, 2016).

User Engagement. User engagement, i.e., the attention given by the user to the SAR, is an effective measure of the usefulness and effectiveness of the SAR (Clabaugh et al., 2019). Having shown the benefits of creating a long-term relationship between the human and the SAR in order to motivate the human to use the SAR, Kidd and Breazeal (2008) stressed the importance of user engagement during the interaction to allow said relationship to be formed, gain the human's trust and, consequently, assist the user on a daily basis. Given the aforementioned difficulty in retaining the attention of the users for periods of time long enough to enable long term interactions, tackling this issue is crucial for social assistive robotics (Leite et al., 2013).

In human-SAR interactions, user engagement is produced by behavioural, affective and cognitive aspects (Clabaugh et al., 2019). Therefore, any approach that promotes the development of one of these elements can strengthen the interest of the user in the robot.

According to Leite et al. (2013), employing approaches that enable SARs to display social behaviours is essential to maintain user engagement. For example, the perception that the behaviour of the user affects and is affected by the behaviour of the SAR leads humans to recognise the robot as human and as a social agent. Strategies that portray the SAR as a human and a social agent set the basis for emotional and social interactions. Consequently, they contribute to the ability of the SAR to draw and retain the attention and availability of the users (Leite et al., 2013; Rodríguez-Lera et al., 2018).

Simulating physical abilities such as making eye contact (Kidd & Breazeal, 2008), exhibiting facial expressions (Clabaugh et al., 2019) and subtle anthropomorphic behaviours (e.g., eye blinking and pupil dilation) (Leite et al., 2013; Scassellati et al., 2018), and displaying mechanism of joint attention (Scassellati et al., 2018) can contribute to make robots appear more human and smart. Moreover, selecting an exaggeratedly complex appearance that does not reflect the social capabilities of the SAR can provoke unrealistic expectations from the user. These expectations boost the novelty effect but, once the effect dissipates, they cannot be met by the behaviour of the SAR. Based on these premises and on experimental results, prior research substantiates that choosing a suitable physical appearance that reflects the social skills of the SAR is crucial to retain the attention of the users (Leite et al., 2013). Both the implementation of the aforementioned physical abilities and the choice of a suitable appearance for the SAR help the development of engaging interactions. Consequently, these two aspects are important to long-term interactions and are commonly taking into consideration in literature. For example, Clabaugh et al. (2019), Kidd and Breazeal (2008), and Scassellati et al. (2018) exploited the aforementioned non-verbal techniques to enable the SAR to display a human and social behaviour.

Apart from the physical skills that were already described, many authors proposed and implemented techniques based on social processes rather than on physical ones. These social processes can lead the SARs to display even more complex social and intelligent behaviour. Examples of these social processes include techniques based on verbal communication (Leite et al., 2013), adapting the behaviour of the SAR based on its relationship with the user and on the circumstances of each interaction (Kidd & Breazeal, 2008), personalising the SAR to the user (Clabaugh et al., 2019; Scassellati et al., 2018; Tapus & Mataric, 2008), and avoiding repetitive behaviour (Rodríguez-Lera et al., 2018).

There are several techniques that can be used during verbal communication to induce social and personal behaviours and, consequently, engagement. Starting and finishing interactions with proper greetings and farewell expressions (respectively), generating interactive dialogues rather than monologues, adjusting conversational content based on previous encounters, and employing mechanisms of self-disclosure have proven to be effective tools to extend user engagement. A SAR is said to possess mechanisms of self-disclosure if it reveals personal information, opinions, or their back-story to the users as their relationship develops, similarly to what occurs in human relationships. Moreover, when it comes to long-term interactions with children, the complexity of the behaviour and the quantity of cues displayed by the SAR contributes towards maintaining the interest of the children (Leite et al., 2013).

Estimating a Relationship. The SAR developed by Kidd and Breazeal (2008), which aimed at accompanying users throughout 6 weeks in order to help them lose weight, tracked the status of its relationship with the user and acted accordingly to it. As discussed in Section 4.2, the results demonstrated that the relationship estimator and the script generator (which depended on the estimator's output) lead the users to develop a relationship with the SAR and prompted the users to engage with this weight loss method for longer and more often than with traditional methods (Kidd & Breazeal, 2008). Furthermore, Kidd and Breazeal (2008) stated that it is crucial to define the type of relationship that the SAR will try to create with the user based on the intended application of the SAR. For example, in this study in particular, the SAR attempted to generate a relationship based on support and care given that the SAR aimed at helping the user to lose weight. The SAR generated this type of relationship by using a nurturing and positive tone and kind expressions. According to Kidd and Breazeal (2008), defining the type of relationship between the user and the SAR is important to design a system that can robustly and coherently control the relationship status.

When SARs interact with several users, the capability of recognising repeated users is an added benefit. Moreover, creating relationships with users and acting with each one of them based on the corresponding relationship status contributes to simulating a more realistic social behaviour (Leite et al., 2013).

Simulating a personality. The tone and the style of communication used by the SAR developed by Kidd and Breazeal (2008) can be regarded as an attempt to simulate a personality in the robot. This is another approach that has been used to lead users into perceiving the SAR as a human. An individual's personality comprises an estimate of the emotional, personal, and social characteristics that are frequently and consistently displayed by that individual. Thus, personality shapes the social interactions between humans. Additionally, personality and behaviour are believed to be intrinsically entangled. Therefore, understanding the personality of an individual can allow a rational agent to know which behaviours to expect from the individual (Tapus & Mataric, 2008).

Kidd and Breazeal (2008) designed the SAR to display a caring and supportive personality, which contributed to increase user engagement. Nevertheless, Tapus and Mataric (2008) demonstrated that, while this type of personality is effective for some users, a different type can be more effective for other users. Therefore, Tapus and Mataric (2008) enabled the SAR to display a personality based on two parameters that were adapted to each user, rather than displaying the same personality to every user. Although Tapus and Mataric (2008) was a short-term interaction study, the implementation of a personality that was adapted to the users yielded important results for long-term interactions.

Personalisation. The users assisted by SARs have different needs, preferences and goals, and belong to different social and age groups. Moreover, the learning, recovery or development curve of each person has different characteristics (Clabaugh et al., 2019; Clabaugh & Matarić, 2018). Consequently, each user requires a different optimal assistive approach that addresses their unique personality (Clabaugh et al., 2019; Tapus & Mataric, 2008). Given the inherent goal of helping the user in the most effective way, SARs must adapt their coaching, motivation and socialisation techniques to each user's personality, regardless the task in which the robot must assist them with (Tapus & Mataric, 2008). These strategies are known as personalisation. When assisting children with ASD, personalisation is even more crucial: given that users with this disorder present very specific and unique characteristics, shaping the assistive methods to each one of them is important to ensure an optimal learning procedure (Scassellati et al., 2018). The work done by Clabaugh et al. (2019) shows the importance of personalising a home-in SAR for children with ASD. The authors considered the long-term duration of their study to be essential to achieve a successful personalisation, since the personalisation parameters only converged after several sessions.

Nevertheless, personalisation strategies carry challenges. Getting to know a person and their traits requires several interactions. Similarly to what happens with humans, in short-term interactions, SARs cannot get a representative idea of someone's personality. Consequently, they cannot adapt their behaviour accordingly to the personality of the user in short-term interactions. Contrarily, long-term interactions not only provide enough data to enable personalisation, but also require it (Clabaugh et al., 2019; Clabaugh & Matarić, 2018; Leite et al., 2013). After gathering the necessary information about the users, it is necessary to process that information in order to respond to the complex needs of every individual. This process brings many underlying obstacles such as the variety of the needs and the preferences of the users, as well as data noise. (Clabaugh & Matarić, 2018). In order to address them, the majority of the studies found in literature that implemented personalisation resorted to computational personalisation (Leite et al., 2013), i.e. employing machine learning techniques that autonomously learn the behaviours that are the most suitable for every user (Clabaugh et al., 2019).

Several studies were able to personalise SARs to their users, and demonstrated the importance of personalisation. For example, Tapus and Mataric (2008) adapted the robot's personality traits to every user, and Clabaugh et al. (2019) and Scassellati et al. (2018) adjusted the challenges and games to every child. Estimating the evolution of the relationship with a user and acting according to it, as performed by (Kidd & Breazeal, 2008), can also be considered a type of personalisation. Moreover, the SAR that was designed in this study was capable of engaging in small talk with the user. Said small talk was personalised to users according to variables such as personal data, the duration of the last interaction, and the relationship status. Additionally, the SAR was capable of exhibiting memory by greeting the user differently depending on whether the interaction was the first interaction of the day or not. These techniques contributed to maintain user engagement.

Adaptive behaviour. Finally, adaptive behaviour has also been employed in long-term interactions. In human-computer interaction, adaptiveness consist in the "dynamic adaptation by the system itself to current task and current user" (Fischer, 2001). In the SAR domain, adaptive behaviour can be regarded as the ability of the SAR to change its conduct to accommodate certain changes in the environment, in the task or in the user. According to Tapus and Mataric (2008), the adaptive behaviour of an SAR must address both short- and long-term changes. Given the unpredictable scenarios in which SARs are expected to assist their users, it is of utmost importance that these robots are adaptive (Matarić & Scassellati, 2016). For example, when the SAR has the objective to assist the user at home, it is crucial that it can adapt its behaviour to the different scenarios which can occur in such an uncontrolled environment (Rodríguez-Lera et al., 2018). Moreover, in the study carried out by Clabaugh et al. (2019), the answers of the users to a questionnaire demonstrated that the usefulness of the SAR depends on its adaptability, since the users who considered the robot more adaptive also considered it more useful.

Several techniques to promote adaptive behaviour have been employed in the projects that were analysed in the present literature survey. For example, Rodríguez-Lera et al. (2018) proposed a control architecture capable of adapting the conduct of the robot to external factors such as the number of strangers surrounding the robot, or the lack of navigation information. Additionally, the SAR designed by Kidd and Breazeal (2008) adapted its verbal communication to the current state of the user (based on input inserted by the user) and time of the day. Additionally, personalisation can be regarded as a specific type of adaptive behaviour, since it provides SAR with the tools to change their conduct to best suit the user and their characteristics (Clabaugh et al., 2019; Tapus & Mataric, 2008). Tapus and Mataric (2008) interpreted the physiological signals of the users to adapt the behaviour of the SAR in order to achieve optimal performance from the user. An interesting approach that could have been taken was the interpretation of the physiological signals of the user to adapt the behaviour of the SAR to the inner state of the user in each session, throughout different sessions. Despite recognising the need to respond to long-term changes as well as short-term ones, Tapus and Mataric (2008) only comprised one short-term session with each user, as it was already mentioned. Therefore, in this study, it was not possible to adapt the actions of the SAR based on the physiological cues from the same user in different sessions, given that there was only one session.

Several studies have been able to respond to subtle changes in the environment and user. Nonetheless, most SARs and their control systems that can be found in the literature were designed for one specific purpose. Therefore, they do not have the capability to act realistically in unexpected circumstances. Even when they possess mechanisms that aim at imitating social behaviours, these are usually tailored to the one task they were designed for. Thus, the majority of SARs are not able to improvise when faced with slightly different scenarios than the ones they were designed for (Rodríguez-Lera et al., 2018).

In this section the distinction between short-term and long-term interactions was analysed. Although short-term interactions are more common since they carry less challenges and require less resources, long-term interactions yield more insightful and relevant conclusions.

Regarding the minimum duration to distinguish short-term from long-term interactions, Leite et al. (2013) defends that a study should be considered long-term if the user interacts with the SAR for a period of time long enough for the novelty effect to disappear. In other words, if the user has the chance to interact with the robot after no longer being under the initial enchantment caused by the SAR's apparent sophistication, the interaction is considered long-term. When (and if) this point is reached and surpassed, the attention of the user on the SAR depends only on the usefulness and effectiveness of the robot, rather than on a bias caused by the novelty effect. Thus, once the novelty effect disappears, the relationship between the human and the SAR is not expected to suffer alterations that can bias the conclusions taken from the study. Therefore, only by surpassing the time required for the novelty effect to fade away is it possible to analyse the human-SAR relationship and take conclusions regarding how well the robot is expected to fulfil its assistive role.

Sustaining user engagement was found to be the main challenge of long-term interactions. Different techniques that aim at tackling this issue were described. Nonetheless, most studies combined more than one of the mentioned techniques. Humans process the information about others based on data collected from different sources, and employ different types of behavioural cues to express themselves. Similarly, merging the aforementioned different techniques yields a more robust decision-making process. (Clabaugh & Matarić, 2018).

While some of these methods consisted in physical approaches, such as non-verbal cues, others depended on cognitive aspects of human social interactions. Nevertheless, all the analysed studies tackled the challenges of long-term interactions by focusing on the behaviour of the SAR, i.e., on generating approaches that allowed the robot to display complex behaviours. Despite the resources spent on improving the behaviour of the SAR, none of the research projects that were studied attempted to analyse the users or their decision making process.

4.5. Conclusions

This chapter presented an overview of the state-of-the-art interactions between humans and SARs. More specifically, it discussed the control techniques used to model the decision-making of SARs, as well as the characteristics of the current human-SAR interactions, with a special focus on long-term interactions.

SARs assist users from different social and age groups in many domains, and they fulfil the roles of coaches, peers, assistants and companions. Despite the wide range of fields of application of SARs and the diversity of the target users, the tasks suggested to the users in each interaction were of simple creation and management, i.e., the SAR did not propose tasks that were complex to generate or maintain. Moreover, most authors designed controllers specifically tailored to the tasks that the SAR was meant to assist the human with. For example, the described personalisation and adaptiveness techniques were applied to the tasks themselves: specific parameters that shaped the behaviour of the SAR or the complexity of the task were adapted in order to maximise the performance of the user in that task (Clabaugh et al., 2019; Scassellati et al., 2018; Tapus & Mataric, 2008).

Nevertheless, some studies went beyond merely focusing on task performance to adapt the behaviour of the SAR. Such was the case of the work done by Kidd and Breazeal (2008), which took the relationship between the human and the SAR into account, and Tapus and Mataric (2008), who considered the personality traits of the user to control the behaviour of the SAR. However, these studies employed the aforementioned approaches only superficially, without resorting to models that deeply described the human-SAR relationship or the personality of the user. The personality status that was developed by Kidd and Breazeal (2008) was assessed based on the computation of one formula and on the comparison of the value resultant from that formula with one threshold, while Tapus and Mataric (2008) assessed the personality of the user based on one personality trait: the level of extroversion. Every human has a unique personality that can only be accurately described by taking many aspects into account. For example, two people who are introverted can be extremely different from each other. Therefore, the type of social interactions that can enable the SAR to reach its goal more effectively in each case can be distinct. Describing the personality of the users based on one variable does not provide the SAR with the capability to respond to the specific needs of each user. Moreover, as already pointed out in Section 4.3, Tapus and Mataric (2008) displayed empathy by replicating its outcomes, rather than by modelling the user, contrarily to what humans do.

Regarding the control methods, almost all the identified techniques consisted in model-free approaches, being either rule-based or AI. Nonetheless, the state-of-the-art SARs still fail to engage users for the period of time that is necessary to positively impact the quality of life of their users. Most of these SARs can only interact with the users in the context of the task where they are meant to help the users. As a consequence, they lack the social awareness about general social interactions and fail to deal with unknown issues (Rodríguez-Lera et al., 2018). Moreover, although the most recent AI approaches such as RL can deal with extremely complex environments and train on large amounts of data, the complexity of the systems where they thrive are still far away from the complexity inherent to human social interactions (Clabaugh & Matarić, 2018).

Leite et al. (2013) reinforced the need for interactive dialogues, as well as for interactions in which the behaviours of the SAR and of the human influence each other. The authors additionally explain that the current SARs lack the capability to deconstruct and understand social interactions in order to realistically simulate them. The majority of the approaches attempted so far focus on enabling the SARs to display elaborate behaviours, gestures, and verbal content. Although some projects aimed at adapting the behaviour of the SARs based on different types of user input and response, little has been done to actually model the decision making process of the users. Leite et al. (2013) stated that being aware of the affective state of the users is more relevant than displaying an affective state. Moreover, Matarić and Scassellati (2016) declared that SARs must understand the processes behind human interactions and the minds of other agents in order to effectively identify and shape the social aspects of said interactions. Although the literature agrees on the importance of user modelling, no project was found to adapt their conduct based on the cognitive processes of users. Only Rodríguez-Lera et al. (2018) remarked the potential of employing models that take into account human psychology and cognition in order to regulate the behaviour of the robot in the context of Human-Robot Interaction (HRI). Nevertheless, this study implemented the aforementioned concepts to model only the decision making of the robot, and not the decision making of the user.

When humans reason about others, they develop mental models that explain the behaviour of other people as a consequence of their inner states, preferences, external conditions and how these conditions are interiorised by them, given their personalities. It is the fact that humans understand how other people reason, feel and behave that allows them to achieve social goals, build relationships with others, and respond to their needs. Prior research has been tackling the issue of enabling SARs to act as social agents and respond to the needs of the users like humans do, without providing these robots with an essential tool that humans have: a structured model of the cognitive processes behind the behaviour of other people. All in all, socially assistive robotics requires reliable models that describe the deep cognitive and mental processes that motivate the actions of humans. Furthermore, similarly to what humans do when forming a relationship with someone, SARs must be capable of using the information acquired in each interaction to complete that model. Thus, to properly accompany users and understand their progress throughout long-term interactions, these models must comprise both the evolution of the aforementioned processes and the users' traits that are consistently manifested over time. Only such models can provide SARs with the means to behave realistically and naturally in social interactions and, consequently, to retain the attention of users indefinitely in order to assist them in long-term interactions.

5

Cognition models that analyse and reason like humans

Chapter 4 detected research gaps in socially assistive robotics, in particular the lack of mathematical models that represent the cognitive procedures of humans accurately and reliably. Given these research gaps, this chapter presents an analysis of the frameworks employed in the literature to generate models that analyse and reason like humans. Our main aim is to identify state-of-the-art approaches that can be used in the development of a cognitive model that is capable of addressing the challenges faced by SARs.

Section 5.1 presents a description of Theory of Mind, a framework that represents human cognition. It is important to notice that this theory was proposed in the field of neuroscience and, consequently, was not initially developed with the aim of being mathematically and computationally formulated and implemented. Nonetheless, in the field of engineering and computer science, several authors have tried to develop machine-based ToMs. Therefore, while Section 5.1 focuses on explaining the underlying principles of ToM, Section 5.2 analyses one of the most used and successful frameworks for representing ToM mathematically: Bayesian Theory of Mind or simply BToM. In this section, both the theoretical principles behind BToM and the models proposed in the literature based on this framework are discussed. Furthermore, this section includes a theoretical explanation of Bayesian Networks (BNs), a framework based on Bayesian inference that can be used to represent a model with different variables connected with each other and is potentially interesting for the current research project.

Additionally, in Section 5.3, a model-free approach termed *meta-learning* is explained. Although it was concluded in the previous section that it is necessary to model the human decision making process, given that meta-learning was previously used to build a ToM, the suitability of employing meta-learning in the present research project is assessed. Thus, the method employed in the study that generated a ToM by employing meta-learning, as well as the circumstances in which it was tested, are analysed in Section 5.3.

Finally, several researches have employed Markov Decision Processes (MDPs) and, particularly, Partially Observable Markov Decision Processes (POMDPs) to implement ToM and to represent the human behaviour. Therefore, the projects in which these frameworks were used, as well as the reason why they are suited to represent human cognition, are analysed in Section 5.4.

5.1. Theory of Mind

As mentioned in Chapter 4, ToM corresponds to the ability demonstrated by humans (and other rational agents) to analyse and understand the cognitive procedures of other people or rational agents (Baron-Cohen et al., 1985). According to this theory, the rational agents reason about themselves and others based on inner mental states, i.e. the underlying states that motivate human actions and, consequently, can be externalised through behaviour. A rational agent's beliefs, goals, desires, and emotions are all examples of the aforementioned mental inner states (Scassellati, 2002; Wellman et al., 2001). It is believed that children develop a ToM while they grow rather than being born with it, since the experiments that test certain concepts inherent to ToM (e.g. the Sally-Anne experiment which will be explained later on) are successfully passed by children within a certain age gap, but not by younger children (Scassellati, 2002; Wellman et al., 2001).

In order to have a ToM, a rational agent must be capable of recognising that other rational agents can hold *false beliefs* (Baron-Cohen et al., 1985; Rabinowitz et al., 2018; Scassellati, 2002). Given that rational agents may not be able to observe the entire universe simultaneously (i.e., the observation capabilities of rational agents in general are limited) and that certain parts of the environment may be prone to changes while the rational agents are not observing them, these rational agents can hold beliefs about the environment that do not correspond to the true state of the environment (Baron-Cohen et al., 1985). Therefore, when reasoning about another person or rational agent within the ToM framework, rational agents must acknowledge that others may base their actions not on the real state of the environment, but rather on the their own beliefs (Wellman et al., 2001). This concept is essential for ToM since it sets the distinction between the world and a mental perception of that world. Consequently, this concept is used by researchers to evaluate whether agents (children, animals, machines) possess a ToM (Wellman et al., 2001).

To investigate whether children with ASD have a ToM, Baron-Cohen et al. (1985) created the Sally-Anne experiment, which illustrates the concept of false beliefs. In this experiment, Sally and Anne are in a room with a basket and a box. Sally places an object in the basket and leaves the room. While she is gone, Anne changes the location of the object, placing it in the box. Subsequently, the observer is asked where Sally will look for the object when she comes back. An observer who is capable of acknowledging false beliefs will correctly predict that Sally will search in the basket, while an observer who is not capable of acknowledging false beliefs in rational agents will expect Sally to look for the object in its real location, i.e., the box.

Finally, rational agents are believed to act intentionally based on their beliefs and given their desires and goals (Dennett, 1987). This premise sets the basis for generating artificial representations of ToM.

Although ToM was initially proposed in the field of neuroscience, Scassellati (2002) proposed the creation of virtual ToM models. The author suggested the adaptation of two prior ToM models, which were developed within the scope of neuroscience, in order to generate models that were suitable for robotic implementations. These models were constructed in a hierarchical way, based on lower-level modules that described simpler concepts inherent to ToM. The modules that explained the most complex reasoning processes, such as reasoning about the mental states of an observed rational agent and the representation of these mental states, were represented in a higher hierarchy. The lower-level modules controlled low-level visual, tactile and auditory inputs, as well as motor skills. Although the models that were analysed by Scassellati (2002) were not ideal for robot implementation, since the interaction between the different modules was not clearly specified, their structure was deemed a suitable starting point. Therefore, Scassellati (2002) suggested a mathematical formulation of these models that addressed the links between the modules and was suitable for a robotic implementation.

Later on, more authors have proposed and implemented ToM models. The most relevant works in this area, as well as the frameworks used to achieve them, are described in the following sections.

5.2. Bayesian Theory of Mind

Humans intuitively process the world in terms of probabilities, i.e., they make connections between different premises in order to estimate the likelihood of potential outcomes based on prior knowledge (Johnson-Laird, 1994). Furthermore, rational agents reason about the mental states and decision making process of other rational agents according to the beliefs, goals, personality, preferences, and other inner mental states of the observed rational agents (Jara-Ettinger et al., 2016). Therefore, researchers widely agree that the process of analysing and understanding the behaviour and decision making of other rational agents an be formulated within a Bayesian framework (Baker et al., 2007; Jara-Ettinger et al., 2016; Saxe & Houlihan, 2017).

Bayesian inference estimates the probability of occurrence of a random event for a certain system, based on the Bayes' theorem given by the following equation:

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)} \tag{5.1}$$

The Bayes' theorem describes the conditional probability of a random event A, using previous knowledge about random events (for example, event B) that may be related to event A. In other words, in case prior knowledge corresponding to observed events exists for the system, by integrating the information yielded by these observations, it is possible to infer information regarding new random events or random variables for which no observation exists (yet) (Dempster, 2008).

Describing the behaviour of the rational agents in probabilistic terms and as a consequence of the likelihood of their inner states and the surrounding environment, as well as integrating the observed behaviour of rational agents with the knowledge from past observations, sets the foundation for BToM.

Prior studies recognise that ToM can be described based on the principle of rational action, which states that humans act approximately rationally towards maximising the fulfilment of their goals, given their beliefs about the external world (Baker et al., 2007; Dennett, 1987; Jara-Ettinger et al., 2016; Saxe & Houlihan, 2017). Formally speaking, predicting the actions of logical agents based on their inferred mental states is called *forward inference*, while inferring the inner mental states of rational agents according to their observed actions is called *inverse inference*. This terms, which were introduced by Saxe and Houlihan (2017), will be used throughout this chapter and in the remaining document.

Based on the principle of rational action, Baker et al. (2007) proposed three BToM models of observant rational agents (called M1, M2 and M3) to infer about the goals of observed rational agents given their actions. In order to model the cognition of the observant agents, Baker et al. (2007) expressed the action planning of observed rational agents as a function of their goals and the world conditions using an MDP, in which the reward function depended on the goals of the rational agents. Since the MDP framework states that agents choose their actions with the goal of maximising their cumulative reward, this implementation yields a mathematical formalisation of the principle of rational action. Nonetheless, it was assumed that the transition function from a mental state to the subsequent mental state for the rational agent is deterministic, i.e., actions always have the intended outcome. The policy was computed using value interaction. In order to infer about the goals of the observed rational agents based on their behaviour, each one of the three proposed MDP-based model was inverted using *Bayesian inference*. When applying Bayesian inference to the models, prior knowledge concerning potential goals was encompassed. Particularly, the three MDPs models were used to infer about the goals of rational agents moving in a two-dimensions environment, where these goals included particular locations in the two-dimensions world. The behaviour of the observed agents was represented by simple *stimuli* displaying paths between locations (see Figure 5.1). To access the accuracy of the inferences made by the three models, human participants also observed the same stimuli and made their own inferences. The accuracy of the inferences was evaluated by measuring how similar the predictions made by M1, M2 and M3 were to the predictions made by the humans, since the goal of this research project was to develop a model that reasons like humans. The first model (M1) assumed that the logical agents had one constant goal. The second model (M2) assumed that the logical agents had sub-goals, where the number of sub-goals was controlled by a parameter. M3 assumed that the goal(s) of the logical agents may change over time, where the frequency of this change was controlled by a parameters. Several M2 and M3 model variants were generated with different values for the parameters that defined them. For both M2 and M3, the parameters that yielded the model variant whose inferences were the most similar to the ones made by the humans were selected. Finally, the M2 and M3 models generated with the selected parameters were analysed in the conclusions. Baker et al. (2007) concluded that only M2 and M3 were able to accurately predict stimuli where the observed agents took indirect paths i.e., paths that considered an intermediate point between the starting point and the final point. This result reveals the importance of considering non-static goals. Regarding the M2 model, there was the risk to include an excessive amount of sub-goals with the objective to explain all the detours from the main path, even the deviations caused by noise. To avoid this, Baker et al. (2007) stated that the sub-goals must only be comprised in the model when, considering different initial conditions, it has been observed that sub-goals have been always considered by rational agents in the past. In these experiments, the initial conditions corresponded to the starting point of the observed agent, the position of the barrier and the position of the potential goals. Moreover, considering dynamic goals that may be replaced over time by new goals requires encompassing a factor that determines the frequency of the changes in the goals in order to accurately capture the human behaviour, since humans do not change their goals unless they have a reason to do so.

Later on, Baker et al. (2008) focused on the importance of social goals in real-life interactions of humans. Therefore, in order to develop a model that incorporates the context of social interactions, the possibility of rational agents making inferences about social goals of other rational agents was added to the model that was proposed by Baker et al. (2007). The authors argue that processes based on inter-

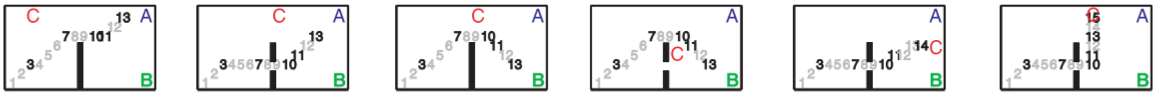


Figure 5.1: Examples of the stimuli used by Baker et al. (2007) to represent the behaviour of the observed rational agents. These stimuli were analysed both by humans and by the rational observant agent developed by the authors. Source: Baker et al. (2007)

preting exclusively physical cues (for instance, measuring the evolution of the distance between rational agents in order to determine whether one rational agent is chasing the other one) are not comprehensive enough to explain complex social inferences, since they lack the environmental and situational context. Thus, a similar method as the one proposed by Baker et al. (2007) was implemented, in which the action planning of rational agents was described by an MDP and was inverted using Bayesian inference. Considering experiments of agents moving in two-dimensional environments, the models that included second order beliefs, i.e., the belief of rational agents about the goals of other rational agents, yielded the most accurate results in terms of imitation of human reasoning. The achieved results stress the importance of the "ability to represent a mind that represents another mind" in order to capture human behaviour in social context.

Nonetheless, these two studies (Baker et al., 2007; Baker et al., 2008) assumed that rational agents have integral access to the surrounding environment while, in reality, rational agents are only capable of observing a part of the environment. Hence, the aforementioned assumption prevented the rational agents to recognise that other rational agents can hold false beliefs, which prior studies widely agree that is an essential characteristic to possess a ToM (Rabinowitz et al., 2018; Wellman et al., 2001), as explained in Section 5.1. To tackle this problem, in a later study, Baker et al. (2011) expanded the previously proposed model in order to encompass the *principle of rational belief*, which states that beliefs of rational agents result from the combination of external conditions and the perceptual access that the rational agents have to those conditions. In order to do so, Baker et al. (2011) used POMDPs to implement their model. POMDPs encompass the main principles of MDPs, while incorporating the premise that rational agents are only able to observe a part of the environment at each time, and thus they should hold a belief of the entire state space that is described in probabilistic terms. By applying Bayesian inference to the POMDP, the authors managed to jointly infer the beliefs and goals of rational agents that moved in two-dimensional spaces, based on their actions. The estimation of beliefs and goals encompassed their level of strength for the rational agent. The inferences yielded by the BToM that jointly inferred the beliefs and goals and that explicitly took into account the rational agents' observations were very similar to the ones reached by humans. These results demonstrated that it is crucial to jointly infer the goals and beliefs and to encompass the observation process of rational agent in order to achieve an analysis and reasoning approach that closely resembles that of humans.

Finally, Baker (2012) proposed the model represented in Figure 5.2, in which the beliefs and goals of rational agents are additionally influenced by, respectively, *general world knowledge* and *general preferences* of the rational agent. These are described to be "high-level variables that apply across situations" and are both inherent and specific to each rational agent. Therefore, these concepts lay the foundation to expand such a cognition model to allow for personalisation according to different individuals. Nonetheless, since Baker et al. (2011) merely focused on short-term interactions of rational agents, the two high-level variables mentioned above were excluded from their experiments, since these two variables could only be inferred throughout several sessions, in long-term interactions. Thus, only the variables displayed inside the situation box, as represented in Figure 5.2, were contemplated in the model employed in the experiments.

All in all, the models described above managed to accurately simulate the analysis and reasoning procedures of humans when they infer the beliefs and goals of other rational agents, in both individual and social circumstances. Nevertheless, these researches did not address the application of their cognitive models to scenarios that simulate representative and realistic human interactions, as all the experiments were conducted in simple, two-dimensional environments. Additionally, the models that were developed to predict the beliefs of the rational agents acted as merely observers, i.e. these models did not interact with the rational agents whose mental states were being inferred: instead, rational agents performed actions and their behaviour was presented both to humans and to the models, who made estimates. Contrarily, given the scope presented in Chapter 4, in the present research project

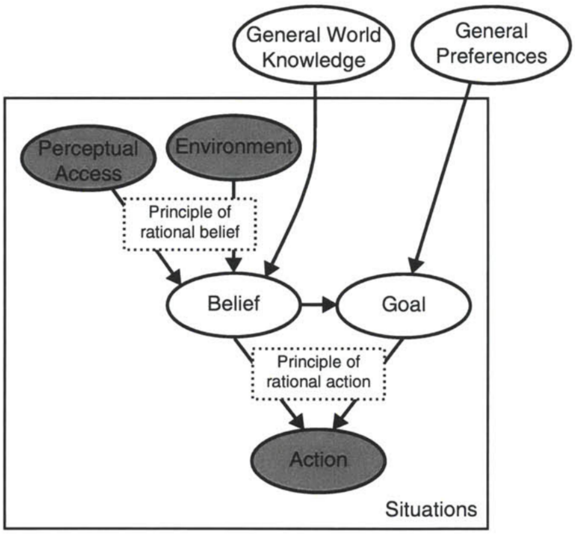


Figure 5.2: Cognitive model of a rational agent proposed by Baker (2012). This model explains the behaviour of a rational agent based on the principle of rational action (i.e., intentional actions of the rational agent are assumed to be influenced by the rational agent's beliefs and goals, where the rational agent aims at optimising the fulfilment of its beliefs and goals upon its actions), the principle of rational belief (i.e., beliefs and goals of a rational agent result from both the surrounding environment and the perceptual access that the rational agent has to its environment), and the influence of general world knowledge and general preferences on, respectively, beliefs and goals of the rational agent. Source: Baker (2012)

the rational agent must interact with human agents while estimating their mental states.

Despite the pitfalls that were discussed, the models that were explained should be considered as starting points for future research since they demonstrate a strong potential to grasp ToM. Furthermore, since they can be expanded to more complex environments (for example, by including the aforementioned high-level variables), these studies can set the basis to model the analysis, reasoning, and decision making of rational agents in real-life situations.

Similarly to the inner mental states considered in most formulations of BToM (such as beliefs and goals), emotions play an important role in determining the decisions and actions of rational agents. Noticing this, Saxe and Houlihan (2017) remarked that including emotions in the models that explain how humans reason about the cognitive procedure of other rational agents is essential to accurately describe this process. Henceforth, the authors proposed a framework to accurately estimate a wide range of emotions, using a Bayesian Hierarchical generative model.

Predicting emotions based only on behaviour, physiological cues, and body language leads to ambiguous conclusions regarding the emotions of logical agents, since these manifestations can be the same even when triggered by different emotions. Thus, Saxe and Houlihan (2017) proposed to use the environment of the logical agent to address the aforementioned ambiguity.

Furthermore, since emotions are unique and particular, predicting them based on actions (inverse inference) yields highly specific but usually biased estimations. Contrarily, predicting actions from emotional states (forward inference) is generally accurate, but the resultant predictions are very general, i.e., they include the generic type of action that the rational agent will take but with low specificity. Therefore, Saxe and Houlihan (2017) suggested to combine both processes: firstly, predicting a specific, particular emotion based on the observed actions (inverse inference); secondly, predicting upcoming behaviour of the rational agent based on the estimated emotion (forward inference). Finally, one can confirm or contradict the emotion that was originally predicted, through comparing the observed/realised behaviour with the behaviour that was predicted accordingly. In case needed, the model can be adjusted based on this assessment. This framework differs from previous theories and approaches that were suggested for emotion estimation, in that it attempts to merge two distinct approaches in order to infer the (interrelated) behaviours and emotions of logical agents, while previous researches focused on estimation or prediction of only one. Although the study carried by Saxe and Houlihan (2017) is merely conceptual and no formal model was described, developed or implemented, the suggestions provided in their research are highly relevant, and provide a potential to extend the models and approaches that were previously developed to fit realistic human interaction contexts, rather than just applying these models and approaches to simple navigation tasks. Moreover, the concept of combining inverse inference to predict mental states (which yields rather unique predictions) and applying forward inference to correct the prior prediction is a technique that can be applied to other mental states besides emotions.

In the more recent years, several studies have applied BToM to more realistic scenarios. For example, J. J. Lee et al. (2019) studied human-robot nonverbal interactions between a speaker and a listener in a story-telling context. The cognitive model of both logical agents were developed according to BToM. On the one hand, the goal of the speaker was to make an inference about whether or not the listener is attentive by decoding the cueing behaviour of the listener, and, if needed to, change the status of attention of the listener by resorting to non-verbal speaker cues. On the other hand, the listener aimed to infer the speaker's belief regarding its own attention state and to manipulate it, in order to induce the speaker to believe that the listener was attentive.

The reasoning and decision making procedures of the speaker were modelled using a POMDP in which the states corresponded to the possible attentive states of the listener, the actions corresponded to the possible speaker cues and the observations corresponded to the listener's responses. Including the cost of the actions by the speaker was deemed important in order to prevent the speaker from constantly cueing the listener, which would be ineffective. J. J. Lee et al. (2019) used a soft-max policy. Nonetheless, there was little information regarding the parameters that should be used in the functions of the POMDP (the transition, observation and reward function). Thus, a technique called *Apprenticeship Learning* (Makino & Takeuchi, 2012) was employed. This technique yields the parameters of the POMDPs in environments where little information about the model is available, based on the demonstrations performed by human experts. It is assumed that the expert represents the POMDP with the parameters that are optimal to fulfil the task and that the human aims at maximising the reward. The parameters that lead to a optimal policy that generates a behaviour more similar to the behaviour of the human expert, are chosen. In this case, tuples of *state, observation, and action* were collected from human experts who demonstrated the task of speaking while making the listener to be attentive. The parameters that resulted from the apprenticeship learning were used in the training and yielded the policy that was used by the speaker.

The reasoning and decision making procedures of the listener were modelled as a Dynamic Bayesian Network (DBN). In this network, the state corresponded to the status of the speaker's belief, while the actions and observations were, respectively, the observations and actions of the speaker's POMDP model. The selected policy consisted in choosing the action that yielded the best immediate belief.

The POMDP model of the speaker was compared with a Hidden Markov Model and a Multivariate Hidden Markov Model in terms of how well it approximated the tuples of states, observations, and actions of the speaker. Contrarily to the POMDP, the Hidden Markov Model and the Multivariate Hidden Markov Model did not consider the effect of the speaker's actions on the listener's state or behaviour. The proposed POMDP performed slightly better than the other two models. This result highlights the importance of these two aspects of POMDPs. The efficacy of the DBN model corresponding to the listener was evaluated in experiments that consisted in interactions of the listener with children. The listener modelled by a DBN performed better than when a model was used that did not infer about the speaker's beliefs in order to produce the listener's cues, where these results support the conclusions reached by Baker et al. (2008) regarding the importance of representing second order beliefs in social interactions.

Finally, a limitation was identified in the work by J. J. Lee et al. (2019) regarding the definition of the cost function. More specifically, a cost function that was formulated according to each listener could have yielded better results, since different individuals display unique behaviours as listeners. To tackle this shortcoming, it was suggested to adapt the costs to each individual in the reward function. However, that required performing long-term interactions, where both long-term interactions and penalisation procedures were out of the scope of J. J. Lee et al. (2019). Despite not being implemented, personalisation of the costs in the reward functions is valuable for works that focus on long-term and realistic interactions, which is the case of the present study.

5.2.1. Bayesian Networks

Bayesian Network (BN) is a framework based on Bayesian inference that corresponds to a graphical representation of different random variables that compose a probabilistic system, as well as the causal relationships between them, where these relationships are defined via conditional probabilistic terms (Heckerman, 2008).

The structure of a BN corresponds to a Direct Acyclic Graph (DAG), which displays the links between a set of n variables $\mathbf{X} = \{X_1, ..., X_n\}$. Each variable is represented by a *node*. The nodes that influence node X_i are called *the parent nodes of* X_i, written as $\mathrm{Par}(x_i)$, and x_i is a child node for each of these parent nodes. The influence of a parent node over a child node is graphically represented by an *arc* and is comprised in the mathematical model through the conditional probability $p(x_i|\mathrm{Par}(x_i))$. Nonetheless, while an arc represents the influence of one node over the other, the conditional probability $p(x_i|\mathrm{Par}(x_i))$ concurrently encompasses the influences of all the parent nodes of X_i over that child node X_i.

The joint probability of the network is based on Bayes' Theorem (Equation 5.1) and given by (Heckerman, 2008):

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\mathrm{Par}(x_i)) \tag{5.2}$$

The probabilistic aspect of BNs, which expresses the causal relationships between the random variables, allows to encode the relative strength of these relationships (Heckerman, 2008)

A Bayesian network can be used to represent the inner mental states of humans (which are random variables), the relationships between these states and how they influence the behaviour and decisions of humans. To better represent the relationships between those states, it is beneficial to have a weight associated to each node, since different parent nodes can influence their child nodes with different strength. In the field of computer vision, the work done by Zhou and Huang, 2006 proposed an adaptation of Bayesian Networks in which the edges are weighted. Consequently, the conditional probability can be replaced by an estimation, since the final probability can be tuned using the weights, which are then learnt from training samples. The authors state that the *weighted Bayesian Networks* are particularly useful for cases in which the conditional probabilities cannot be modelled analytically, i.e. when "the likelihoods can only be approximated", which is the case of cognitive processes. The underlying principle behind this implementation is that "the conditional probabilities are weighted by their relative importance in determining the posterior probability." However, when the mathematical model is formalised, the weights are defined for each node, not for each arc. Thus, each weight affects the conditional probabilities of all parent nodes over a child node in the same way. Therefore, by using this technique, the weights cannot represent a higher relative influence of a parent node over another parent node on the child node (Zhou & Huang, 2006). Thus, this particular adaption of BN does not bring any additional benefit to the representation of cognitive processes.

One of the approaches employed by J. J. Lee et al. (2019), Dynamic Bayesian Network (DBN), is a particular case of the more generic framework Bayesian Networks. DBNs consist in graphical representations of systems that can be described as BNs and whose variables influence each other in a subsequent moment in time. By representing causal unidirectional relationships between variables in adjacent time steps, DBNs encode the concept that events in a certain moment can trigger events in the future, but not the other way around (Ghahramani, 1998). In spite of the successful implementation of a listener rational agent that could simulate attention in children (see Section 5.2), based on the estimated belief state of the speaker, J. J. Lee et al. (2019) did not motivate the choice of a DBN to model the listener logical agent over other possible approaches (such as the POMDP that was used in the same research to model the speaker rational agent).

Despite the high correlation between the structure of human reasoning and behaviour and the structure of BNs, only a few prior research projects have applied this type of networks to represent human reasoning, decision making, or behaviour. This may be due to the requirement of defining the conditional probabilities for all random variables explicitly, which is highly challenging in such a complex and unobservable system as the human cognition.

5.3. Meta-Learning

The concept of *meta-learning* is a different approach from the ones analysed so far, since it is a model-free technique. This framework proposes the creation of algorithms that can *learn to learn*, i.e., that continually attempt to perform several tasks, adapting and correcting the way they generalise throughout this process (Thrun & Pratt, 1998).

In order to clearly characterise which algorithms are capable of *learning to learn*, Thrun and Pratt (1998) defined such algorithms in the following way: for a family of tasks, the performance of an algorithm that learns to learn improves gradually according to the experience and with the variability of tasks that have been trained. Contrarily, the algorithms whose performance does not progress with the variability of tasks (but only according to experience) are not considered meta-algorithms. In other words, the fact that other tasks from the same family were previously attempted leads meta-learning algorithms to perform better in an unseen task, since they are able to extrapolate the experience acquired to new tasks. Therefore, meta-learning algorithms transfer and generalise the knowledge acquired in attempting to fulfil a task to new tasks (Thrun & Pratt, 1998). This concept is similar to the learning procedure of humans. People face different challenges and tasks in the course of their lives. As they grow up, they are generally able to achieve new tasks with less effort and better performance. This happens because, when faced with a task, humans do not start anew. Instead, they employ skills and experiences previously learnt in related tasks, combining the approaches that have proven to be successful with other approaches that, given their knowledge, have the potential to be suitable to solve the new task. The meta-learning approach inherent to humans results not from the ability to memorise concepts, movements, or sequences of actions *per se*, but from the capacity to generalise them (Vanschoren, 2019).

Systems that comprise a family of tasks or problems (i.e., a group of different tasks with certain similarities) are suitable and ideal to apply meta-learning to solve those tasks (Vanschoren, 2019). Handwriting and speech recognition, computer vision, and personalised user interfaces are examples of such systems (Thrun & Pratt, 1998). Furthermore, the use of meta-learning allows rational agents to avail the experiences that have been collected in previous tasks, regardless of whether or not these experiences have successfully been completed. Therefore, the rational agent does not need to start an unobserved task anew, as long as a similar one has already been attempted at least once by the rational agent. (Vanschoren, 2019).

Given the aforementioned advantages of meta-learning and its similarity with the learning procedures of humans, the exploitation of this framework to represent human cognition was is expected to be promising, and thus was considered in this literature survey as a topic to be investigated in the available literature. Nevertheless, in the literature, only a few studies (see, e.g., Rabinowitz et al. (2018) and Gao et al. (2019)) applied this framework in the context of ToM or human-SAR interaction.

Rabinowitz et al. (2018) proposed a meta-learning approach to build a machine-based ToM. The authors argued that it is unnecessary to model the cognitive procedures of rational agents in detail, since the reasoning of humans in social interactions mainly relies on high-level models of other rational agents In other words, Rabinowitz et al. (2018) believe that humans mainly reason about the cognitive procedures of other rational agents in terms of mental states and abstractions, without explicitly considering the connections between them. Additionally, it was stated that the efficacy of controllers that aim at mimicking the reasoning and decision making procedures of humans are assessed according to how well these controllers predict and understand the behaviour of other humans, rather than how well they correspond to the cognitive procedure of humans. Therefore, Rabinowitz et al. (2018) considered a model-free approach without accounting explicitly for the cognitive procedures of logical agents. In other words, given that Rabinowitz et al. (2018) believed the underlying process of human cognition to be unimportant compared to the final predicting capabilities of the observant rational agent, they suggested to train a black box.

An environment including several rational agents that interact with the environment was created resorting to POMDPs. Each rational agent had unique preferences and made different observations. Thus, each rational agent was characterised by its own reward and observation functions, as well as their own discount factors. Nonetheless, the rational agents shared the same transition and observation functions, as well as the same state, action and observation space definitions, because the rational agents populated the same environment. Hence, the POMDP framework was used to describe both the environment and the agents. Simultaneously, an observer, modelled by a neural network, aimed to predict the future behaviour of the logical agents. The procedure of predicting the behaviour of logical agents consisted of two sub-procedures: one which was general to the family of agents and another which corresponded to personalisation. The first sub-procedure was generalised to new logical agents, while the second sub-procedure was specific to each logical agent. Rabinowitz et al. (2018) stated that the first sub-procedure included the meta-learning component of the training procedure, since the observant rational agent that was developed in this project learnt to learn how to predict the behaviour of unseen observed logical agents based on generalising the knowledge previously obtained in training. In other words, by attempting to predict the behaviour of the observed rational agents in the training phase, the sub-procedure that was general to the family of agents allowed the observant rational agent to, posteriorly, be able to predict the behaviour of new rational agents. Although the results demonstrated a strong capability to predict both the characteristics that were common to all logical agents as well as the traces that were unique to each rational agent, the simulation environments that were used to assess the proposed approaches were simple and did not represent real-life human interactions. These scenarios were different from and simpler than the ones described in Chapter 4, in which the need to model users in human-SAR long-term interactions was detected. Thus, it is unlikely that a model-free approach like the one presented by Rabinowitz et al. (2018) has the capability to address the problems that motivated the aforementioned user modelling, such as being able to respond to the long-term needs of users or to achieve realistic interactions with SARs.

Given that a model-free meta-learning approach is anticipated to be unsuccessful in the present research, the possibility of selecting a meta-learning model-based technique is the only viable option to use this framework. Taking into account that meta-learning thrives in systems that present several tasks belonging to one family of tasks, such as personalised user interfaces (Thrun & Pratt, 1998), the aforementioned framework can be beneficial when developing a cognitive model that can be adapted to several individuals. In such case, the more individuals that participate in the experiment, the better performance is expected from the rational agent. Therefore, depending on the number of participants considered in the study, it might be advantageous to chose a meta-learning approach.

5.4. The use of Markov Decision Processes (MDPs) to represent human cognition

Baker et al. (2007) demonstrated that MDPs are suitable to represent human behaviour according to the principle of rational action, since the structure of this framework (see Subsection 4.3.1) can express mathematically that rational agents behave according to their goals, desires, and external constraints, given their beliefs about the surrounding world.

Taking into consideration that, according to the principle of rational action, agents tend to choose the actions that maximise their goals given their beliefs about the world, MDP is a suitable approach to model this principle. By including the goals of agents in the reward function and through choosing suitable state and action spaces, MDPs can determine the behaviour of agents according to the principle of rational action. Furthermore, since literature agrees that this principle is an underlying concept of ToM, as mentioned in Section 5.2, MDPs are a promising technique to represent human cognition within the ToM framework.

Nonetheless, MDPs assume that agents have complete access to the external environment, precluding the representation that agents can hold false beliefs. Therefore, modelling rational agents using this framework yields an inaccurate description of human reasoning, as noticed by Baker et al. (2011). Consequently, the concept of Partially Observable Markov Decision Processes (POMDPs) is explained next.

5.4.1. Partially Observable Markov Decision Processes (POMDPs)

In order to address the shortcoming of MDPs in dealing with situations where agents do not have access to or cannot observe all the states of the environment, several studies employed POMDPs rather than MDPs to represent action-planning and behaviour of humans (Baker et al., 2011; J. J. Lee et al., 2019; Pineau & Thrun, 2001).

The use of POMDP by several authors to represent human behaviour was driven by the capability of POMDPs to represent the agent's limited perception of the environment (which constrains the decision making of agents) and the aim of the agents to choose actions that satisfy their goals and desires. Examples of research projects that have used POMDPs to model the analysis and decision making of rational agents are Baker et al. (2011), J. J. Lee et al. (2019), and Rabinowitz et al. (2018) which were explained in detail in Section 5.2 and Section 5.3. Moreover, in some cases, the necessity to deal with uncertain environments contributed to the employment of POMDPs. For example, Pineau and

Thrun (2001) applied POMDPs to human-robot speech dialogues with the goal of being able to deal with unpredictable responses from the human users.

Although POMDPs suit the concepts that command human behaviour, solving a POMDP in complex systems is usually computationally expensive or infeasible (Foka & Trahanias, 2007; Pineau & Thrun, 2001). Various research projects, hence, have proposed several techniques to provide alternative approaches that find approximate solutions of POMDPs, and are computationally more affordable than finding the exact solution. Examples of such alternative approaches include compressing the state or belief state (Foka & Trahanias, 2007).

5.4.2. Hierarchical Partially Observable Markov Decision Processes (HPOMDPs)

One alternative approach to provide computationally affordable solutions for POMDPs with large state and action spaces is the employment of a hierarchical structure (see Figure 5.3) (Foka & Trahanias, 2007; Pineau & Thrun, 2001). This approach allows to reduce both the number of free parameters and to encode prior knowledge in its structure, yielding a simpler learning process (Theocharous et al., 2004) and, consequently, lower computational requirements.

In the study carried by Pineau and Thrun (2001), a robot was in charged of providing different types of information to a user. Apart from resorting to POMDPs to deal with the unpredictability of human behaviour, the authors applied a hierarchical structure to decompose the task in several sub-tasks and, with this, achieve higher computational efficiency.

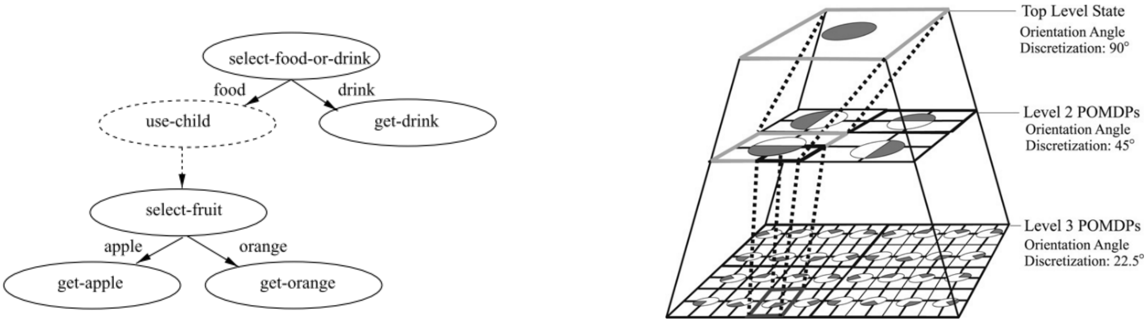
More specifically, the hierarchical structure was implemented by dividing the available informative content into several categories, the management of each being the responsibility of one lower-level POMDP, also referred to as a *child-POMDP* (see Figure 5.3a). The decomposition was achieved by having a recursive structure in which each *parent-POMDP* could select each one of its children. The method of constructing each child-POMDP consisted in removing a subset $\mathcal{S}'$ and a subset $\mathcal{A}'$ from the original process' state space $\mathcal{S}$ and action space $\mathcal{A}$, respectively. The subsets $\mathcal{S}'$ and $\{\mathcal{A}', a_{\text{null}}\}$ corresponded to the child-POMDP's state and action spaces.

Concurrently, on a higher level, a *parent-POMDP* selected the suitable sub-process for each circumstance. Practically, this is conceptually equivalent to choosing the right type of information to be presented. Therefore, the parent-POMDP state space is composed by the same number of states as the number of its children-POMDPs, each one of the states corresponding to one sub-process. Similarly, the action space was made up of the same number of actions as the number of children-processes, each one allowing the selection of one child-POMDP. Despite each child having its own set of actions and states, the observations of all processes were similar, including POMDPs from different hierarchical levels.

Given that each POMDP was individually solved, each process had its own independent optimal policy. Subsequently, the use of this structure allowed to lower the computational effort from more than 24 hours to $480\,\mathrm{s}$, proving the efficiency of the proposed architecture. Nevertheless, the authors stated that the presented structure and method can only be used in systems where groups of actions apply only to certain areas of the state space, since this condition is necessary to perform the described decomposition process.

In the area of robot navigation, Foka and Trahanias (2007) used a Hierarchical Partially Observable Markov Decision Process (HPOMDP) in order to address uncertainties in robot motion. The use of a hierarchical structure was driven by the high dimensionality of the state space and the necessity to achieve high resolution within the mentioned space (see Figure 5.3b). The controller was tested in the real world and the results obtained showed a reduction of complexity and an increment of computational efficiency.

Considering both the suitability of POMDPs to express human behaviour and the complexity of this domain, on the condition that this framework is selected to model human cognition in the current research project, an approach to deal with the complexity of solving the resultant POMDP is likely to be necessary. Given the benefits, the conceptual simplicity and the successful results presented by the studies analysed in this subsection, adopting a hierarchical structure is deemed a good option to tackle the aforementioned issue.



(a) POMDP decomposition carried out by Pineau and Thrun (2001) in order to achieve a hierarchical structure. Source: Pineau and Thrun (2001).

(b) State space decomposition carried out by Foka and Trahanias (2007) in order to achieve a hierarchical structure. Source: Foka and Trahanias (2007).

Figure 5.3: POMDP decompositions employed by Pineau and Thrun (2001) and Foka and Trahanias (2007) to achieve hierarchical structures. In both figures, the highest level of the hierarchy is represented on the top, and the lowest level of the hierarchy is represented on the bottom.

5.5. Conclusions

This chapter presented an overview of the state-of-the-art approaches that represent the procedures of analysis, decision making, and behaviour of humans.

In particular, Section 5.1 it was concluded based on the literature that was reviewed that Theory of Mind (or ToM) is the ideal framework to depict human reasoning. In the field of neuroscience it was demonstrated that the underlying concepts of ToM correspond to how humans analyse, reason about, and understand the cognition and decisions of other humans. It was also discussed that children develop ToM over time as they grow up. Moreover, the studies carried out in the computer science and engineering fields demonstrate the feasibility of applying such a framework to develop models that provide cognition procedures like humans.

Furthermore, as explained in Section 5.2, given the similarity between Bayesian inference and the way that humans process information, ToM can be represented by employing this mathematical theory, yielding Bayesian Theory of Mind (or BToM). Prior applications of BToM were able to generate models that estimate mental state of humans accurately and in a way that resembles the cognitive procedures of humans in real-life. It is important to highlight the work done in Baker et al. (2007), Baker et al. (2008) and Baker et al. (2011), which not only achieved the aforementioned estimations, but also proposed and implemented models that support the addition of more components, such as emotions or variables that define the personality of the observed rational agents. Consequently, the models proposed in these studies can be adapted to more complex scenarios. These models can be used as starting points to develop more elaborated models that can enable personalisation and long-term interactions similar to real-life interactions among humans, which was concluded to be crucial to social assistive robotics in Chapter 4.

Moreover, as explained in Section 5.3, although the concept of meta-learning was deemed interesting because of its conceptual similarity with the learning procedure of humans, applying it in he context of mode ling the human's cognitive procedures, in our opinion, requires a model-based approach. Given that the attempts to use meta-learning to develop a ToM found in the literature were model-free approaches, there was no guarantee that creating a meta-learning model would be feasible or successful in the present domain. Furthermore, given the suitability of the other analysed techniques, it is more reasonable to employ one of the previously mentioned techniques, such as BToM, to formalise the model. However, when personalising the model to the users, it might be useful to consider a meta-learning approach: since these approaches are capable of generalising the learning procedure within different tasks of the same family, the personalisation of the model to each user could be regarded as a different task of the same family. Nevertheless, the decision of adopting a meta-learning approach in this stage should be taken once the model is implemented, based on its characteristics, the type of tasks in which the rational agent will engage and the amount of participants.

Finally, the topics analysed in Sections 5.2.1 and 5.4 present the mathematical approaches that have been more frequently used in the literature to formalise ToM models. On the one hand, Bayesian

Networks (BNs) have the potential to represent the connections between the (random) variables of a system, which can be useful given the complexity inherent to human cognition. On the other hand, Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs) can represent the principles of rationality required for ToM. More specifically, MDPs and POMDPs have proven to successfully represent human's cognition procedures in prior researches. Given the strengths and weaknesses of these approaches, as well as the characteristics of the cognitive model that will be developed, the most ideal approach will correspondingly be selected to formalise and formulate the model.

All in all, the frameworks analysed in this chapter have shown the potential to be employed to develop a model that accurately represents the cognitive procedures of humans. The underlying principles of this model must be defined based on ToM, and the model must be structured considering a BToM implementation. Subsequently, in order to formalise the model, one of the analysed mathematical techniques (BN or POMDPs) must be selected. Finally, although the concept of meta-learning will not be employed to generate, formalise, or implement the model, it might still be employed during the training process.

6
Conclusions

This chapter finalises the literature study and presents the most important outcomes of this survey.

Chapter 4 presented an analysis of the state-of-the-art interactions between humans and SARs. It was detected that socially assistive robotics currently face several challenges that prevent SARs from successfully getting involved in interactions with humans that require deep cognitive-based analysis and prediction of decisions and behaviour of humans. One of the most relevant challenges is the difficulty to engage the attention of the users for long periods of time. Since these robots are not capable of keeping humans interested for long, they are not able to interact with the users consistently and, consequently, cannot assist them as intended. The main reason behind this issue was found to be the lack of social skills and natural behaviour, similar or close to those of humans, that these robots currently demonstrate.

In order to uncover a possible research gap and find a solution to tackle the detected deficiency in the social skills of SARs, an investigation of the control techniques employed to manage the behaviour of these robots was carried out. It was concluded that the majority of the studies found in the literature employed model-free techniques to control the decision making procedures of SARs. Rule-based control systems and AI were the most employed approaches in literature. Furthermore, the majority of the controllers focused on displaying intricate behaviours rather than on understanding the users. Although no study was found to model the cognitive processes of users, some authors stated that SARs must possess an accurate model of their users in order to be capable of interacting with them in a way that realistically resembles human-human interactions (Leite et al., 2013; Matarić & Scassellati, 2016): when humans interact with each other, they develop mental models that explain the behaviours of other humans as a function of their mental states and personality. Additionally, these models are personalised for each individual and, throughout interactions, more and more information about the individual will be included in the model.

Therefore, given how humans interact with each other, SARs lack a reliable model that explains the deep cognitive processes that command human behaviour as a function of their mental states and personality traits, and that allows to estimate these mental states and personality traits. Moreover, this model must evolve as to include the effects of long-term interactions. Only with such a model can these robots be expected to reason and act in a similar way as humans do and get successfully involved in engaging in long-term meaningful interactions with humans. The current project will attempt to develop such a comprehensive cognitive model.

In order to find out about the potential approaches that can be employed to develop the aforementioned model, Chapter 5 comprised a survey of the state-of-the-art approaches used to develop virtual minds that reason like humans. The approach that stood out was according to the Theory of Mind (ToM), since humans are known to analyse and reason about one others' mental and cognitive procedures according to ToM and based on mental states and external conditions.

Considering that ToM is a formalisation of human's cognitive procedures and was not originally developed with the objective to be implemented in computation sciences or robotics fields, BToM was analysed. This technique consists in a mathematical formalisation of ToM and, given the success of prior works that employed this method, it will be used to develop ToM model. More precisely, the model proposed by Baker (2012) will be used as a starting point to develop a model with the capability to respond to the challenges identified in Chapter 4. The selection of this model as a baseline was based on its success in representing human behaviour and on the fact that the structure of this model allows the addition of more elements that can explain the behaviours observed in human-SAR interactions more accurately and comprehensively.

Therefore, the next stage of this research will be the development of a model that can enable SARs to act socially and humanly. For developing a comprehensive and accurate model, it is relevant to investigate and comprise more mental states than the goals and beliefs, which are usually considered by BToM models. Examples of such mental states include emotions and state variables that describe the personality traits of a logical agent. This phase of the research will respond to the first research question (see Section 3.2).

Once the model is complete, it will be mathematically formulated. For this step, the advantages and disadvantages of the techniques presented in Sections 5.2.1 and 5.4 must be evaluated considering the characteristics of the developed model and the sub-questions of the second research question (see Section 3.2). It is important to choose an algorithm with a structure similar to the structure of the model that will be developed. Additionally, the computational efficiency of the two algorithms must be taken into consideration, since the model must be able to run in real time when the human and the SAR are interacting.

Finally, the model must be tested and validated, in order to respond to the third research question (see Section 3.2). The testing procedure aims at identifying the parameters of the model that are suitable for different individuals, as well as evaluating the capability of the model to explain social interactions among humans. The validation phase will focus on using new data to evaluate the model's accuracy. Due to the time constraints imposed by an MSc thesis, and the time consuming nature of carrying user tests, these two phases will probably be carried in a virtual environment.

All in all, this project is predicted to contribute to the field of socially assistive robotics by establishing the importance of user-modelling, since providing SARs with a model of the user's cognition is expected to enable these robots to behave more realistically and naturally. Considering that generating more realistic behaviours is predicted to tackle the very same aspects that are preventing user engagement in long-term interactions and hindering SARs's commercial reliability, this study is expected to bring SARs one step closer to helping users in the real-world.

Bibliography

Baker, C. L. (2012). *Bayesian Theory of Mind : modeling human reasoning about beliefs, desires, goals, and social relations* (Doctoral dissertation). Massachusetts Institute of Technology.

Baker, C. L., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory-based Social Goal Inference. *Proceedings of the Annual Meeting of the Cognitive Science Society*. Washington DC, USA, 1447–1452.

Baker, C. L., Saxe, R. R., & Tenenbaum, J. B. (2011). Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. *Proceedings of the Annual Meeting of the Cognitive Science Society*. Boston, MA, USA, 2469–2474.

Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2007). Goal Inference as Inverse Planning. *Proceedings of the Annual Meeting of the Cognitive Science Society*. Nashville, TN, USA, 779–784.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition*. *21*.(1), 37–46.

Barto, A. G., & Mahadevan, S. (2003). Recent Advances in Hierarchical Reinforcement Learning. *Discrete Event Dynamic Systems: Theory and Applications*. (13), 341–379.

Clabaugh, C., Mahajan, K., Jain, S., Pakkar, R., Becerra, D., Shi, Z., Deng, E., Lee, R., Ragusa, G., & Matarić, M. (2019). Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. *Frontiers in Robotics and AI*. *6*.(110).

Clabaugh, C., & Matarić, M. (2018). Robots for the people, by the people: Personalizing human-machine interaction. *Science Robotics*. *3*.(21).

Dempster, A. P. (2008). A generalization of Bayesian inference. *Studies in Fuzziness and Soft Computing*. *219*, 73–104.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA, USA, MIT Press.

Feil-Seifer, D., & Matarić, M. J. (2005). Defining socially assistive robotics. *Proceedings of the 9th International Conference on Rehabilitation Robotics*. Chicago, IL, USA, 465–468.

Fischer, G. (2001). User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction*. *11*.(1-2), 65–86.

Foka, A., & Trahanias, P. (2007). Real-time hierarchical POMDPs for autonomous robot navigation. *Robotics and Autonomous Systems*. *55*.(7), 561–571.

Gao, Y., Sibirtseva, E., Castellano, G., & Kragic, D. (2019). Fast Adaptation with Meta-Reinforcement Learning for Trust Modelling in Human-Robot Interaction. *IEEE International Conference on Intelligent Robots and Systems*. Macao, China, 305–312.

Ghahramani, Z. (1998). Learning dynamic Bayesian networks. *Adaptive processing of sequences and data structures* (1st ed., pp. 168–197). Berlin, Heidelberg, Germany, Springer, Berlin, Heidelberg.

Hayes-Roth, F. (1985). Rule-based systems. *Communications of the ACM*. *28*.(9), 921–932.

Heckerman, D. (2008). A tutorial on learning with Bayesian networks. *Innovations in bayesian networks: Theory and applications* (1st ed., pp. 33–82). Berlin, Heidelberg, Germany, Springer, Berlin, Heidelberg.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends in Cognitive Sciences*. *20*.(8), 589–604.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*. *50*.(1-3), 189–209.

Jordan, K. S., Pakkar, R., & Mataric, M. J. (2019). Improving Robot Tutoring Interactions Through Help-Seeking Behaviors. *2019 28th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2019*. New Delhi, India.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*. *101*.(1-2), 99–134.

Kahraman, C., Deveci, M., Boltürk, E., & Türk, S. (2020). Fuzzy controlled humanoid robots: A literature review. *Robotics and Autonomous Systems*. *134*.

Kidd, C. D., & Breazeal, C. (2008). Robots at home: Understanding long-term human-robot interaction. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*. Nice, France, 3230–3235.

Lee, C. C. (1990). Fuzzy Logic in Control Systems: Fuzzy Logic Controller—Part I. *IEEE Transactions on Systems, Man and Cybernetics*. *20*.(2), 404–418.

Lee, J. J., Sha, F., & Breazeal, C. (2019). A Bayesian Theory of Mind Approach to Nonverbal Communication. *ACM/IEEE International Conference on Human-Robot Interaction*. Daegu, Korea, 487–496.

Leite, I., Martinho, C., & Paiva, A. (2013). Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*. *5*.(2), 291–308.

Makino, T., & Takeuchi, J. (2012). Apprenticeship Learning for Model Parameters of Partially Observable Environments. *International Conference on Machine Learning (ICML)*. Edinburh, Scotland.

Matarić, M. J., & Scassellati, B. (2016). Socially assistive robotics. *Springer handbook of robotics* (1st ed., pp. 1973–1993). Berlin, Heidelberg, Germany, Springer, Berlin, Heidelberg.

Pineau, J., & Thrun, S. (2001). Hierarchical POMDP Decomposition for A Conversational Robot. *ICML Workshop on Hierarchy and Memory in Reinforcement Learning*. Williamstown, MA, USA.

Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang, C., Eslami, S. M., & Botvinick, M. (2018). Machine theory of mind. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden, 4218–4227.

Rodríguez-Lera, F. J., Matellán-Olivera, V., Conde-González, M. Á., & Martín-Rico, F. (2018). HiMoP: A three-component architecture to create more human-acceptable social-assistive robots. *Cognitive Processing*. *19*.(2), 233–244.

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a Bayesian model of theory of mind. *Current Opinion in Psychology*. *17*, 15–21.

Scassellati, B. (2002). Theory of Mind for a Humanoid Robot. *Autonomous Robots*. *12*.(1), 13–24.

Scassellati, B., Boccanfuso, L., Huang, C. M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P., & Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*. *3*.(21).

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed., Vol. 1). Cambridge, MA, USA, MIT Press.

Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems*. Denver, CO, USA, 1057–1063.

Tapus, A., & Mataric, M. J. (2008). Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. *AAAI Spring Symposium*. Stanford, CA, USA, 133–140.

Theocharous, G., Murphy, K., & Kaelbling, L. P. (2004). Representing hierarchical POMDPs as DBNs for multi-scale robot localization. *Proceedings - IEEE International Conference on Robotics and Automation*. New Orleans, LA, USA, (1), 1045–1051.

Thrun, S., & Pratt, L. (1998). Learning to Learn: Introduction and Overview. *Learning to learn* (1st ed., pp. 3–17). Boston, MA, USA, Springer, Boston, MA.

Vanschoren, J. (2019). Meta-Learning. *Automated machine learning* (1st ed., pp. 35–61). Cham, Switzerland, Springer, Cham.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*. *72*.(3), 655–684.

Zadeh, L. A. (1988). Fuzzy Logic. *Computer*. *21*.(4), 83–93.

Zhou, Y., & Huang, T. S. (2006). Weighted Bayesian Network for visual tracking. *International Conference on Pattern Recognition*. Hong Kong, China, *1*, 523–526.