UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Fábio Grassiotto

CogToM-CST: An implementation of Theory of Mind for the Cognitive Systems Toolkit

CogToM-CST: Uma Implementação Computacional de Teoria da Mente para o Cognitive Systems Toolkit

Campinas

2022

Fábio Grassiotto

CogToM-CST: An implementation of Theory of Mind for the Cognitive Systems Toolkit

CogToM-CST: Uma Implementação Computacional de Teoria da Mente para o Cognitive Systems Toolkit

Dissertation presented to the Faculty of Electrical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Electrical Engineering, in the area of Computer Engineering.

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na Área de Engenharia da Computação.

Supervisor: Profa. Dra. Paula Dornhofer Paro Costa

Este trabalho corresponde à versão final da dissertação defendida pelo aluno Fábio Grassiotto, e orientada pela Profa. Dra. Paula Dornhofer Paro Costa.

Campinas

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

G769c

Grassiotto, Fábio, 1971-
CogToM-CST : an implementation of theory of mind for the cognitive systems toolkit / Fábio Grassiotto. – Campinas, SP : [s.n.], 2022.

Orientador: Paula Dornhofer Paro Costa.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Autismo. 2. Arquitetura cognitiva. 3. Inteligência artificial. I. Costa, Paula Dornhofer Paro, 1978-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: CogToM-CST : uma implementação computacional de teoria da mente para o cognitive systems toolkit
Palavras-chave em inglês:
Autism
Cognitive architecture
Artificial intelligence
Área de concentração: Engenharia de Computação
Titulação: Mestre em Engenharia Elétrica
Banca examinadora:
Paula Dornhofer Paro Costa [Orientador]
Esther Luna Colombini
Ricardo Ribeiro Gudwin
Data de defesa: 01-04-2022
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-1885-842X
- Currículo Lattes do autor: http://lattes.cnpq.br/3571635926402375

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Fábio Grassiotto RA: 890441
Data de defesa: 1 de ABRIL de 2022
Título da Tese: "CogToM-CST: An implementation of Theory of Mind for the Cognitive Systems Toolkit"

Profa. Dra. Paula Dornhofer Paro Costa (Presidente)
Profa. Dra. Esther Luna Colombini
Prof. Dr. Ricardo Ribeiro Gudwin

A Ata de Defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

For Renata and Alice.

Acknowledgements

I must first of all thank my advisor, Prof. Dr. Paula Dornhofer Paro Costa, for accepting a proposal for a work that in many ways is different from what the AI community is looking for, and taking a risk with me. For all the help and suggestions during the process of surveying a difficult subject matter, and proposing something new.

To Profs. Dra. Esther Luna Colombini and Dr. Ricardo Ribeiro Gudwin, for the invaluable suggestions to this work.

To Renata, without whom this work would not have been possible.

To my family, for supporting my choices and motivate me to follow a path of science and technology.

Resumo

Mind-blindness, uma característica típica do autismo, é definida como a falta de capacidade de um indivíduo de atribuir estados mentais a outras pessoas. Essa divergência cognitiva impede a interpretação adequada das intenções e crenças de outros indivíduos em um determinado cenário, resultando tipicamente em problemas de interação social. *Mind-blindness* também pode ser considerada uma característica dos mais sofisticados algoritmos de visão computacional que são capazes de realizar o reconhecimento de padrões em grande escala, mas não interpretam a estrutura causal dos cenários sociais. Neste trabalho, propomos CogToM, uma nova arquitetura cognitiva projetada para processar a saída de sistemas computacionais e raciocinar de acordo com o princípio de Teoria da Mente. Em particular, apresentamos uma implementação computacional para o modelo psicológico de Teoria da Mente proposto por Baron-Cohen e exploramos a utilidade dos conceitos de Affordances e Detecção de Intenção para aumentar a eficácia da arquitetura proposta. Verificamos os resultados avaliando uma falsa-crença típica e uma série de tarefas do conjunto de dados bAbI do Meta.

Palavras-chaves: Autismo; Arquiteturas Cognitivas; Inteligência Artificial.

Abstract

Mind-blindness, a typical trait of autism, is the inability of an individual to attribute mental states to others. This cognitive divergence prevents the proper interpretation of the intentions and the beliefs of other individuals in a given scenario, typically resulting in social interaction problems. Mind-blindness can also be considered a characteristic of most sophisticated computer vision deep learning algorithms that are capable of performing large-scale pattern recognition but still struggle to capture the causal structure of social scenarios. In this work, we propose CogToM, a novel cognitive architecture designed to process the output of computer systems and to reason according to Theory of Mind. In particular, we present a computational implementation for the psychological model of Theory of Mind proposed by Baron-Cohen and we explore the usefulness of the concepts of Affordances and Intention Detection to augment the effectiveness of the proposed architecture. We verify the results by evaluating both a canonical false-belief and a number of the Meta bAbI dataset tasks.

Keywords: Autism; Cognitive Architectures; Artificial Intelligence.

List of Figures

List of Tables

Summary

Preface

This work is born out of a need to understand and help. As a parent of an autistic child, I have followed the footsteps of many just like me. I started trying to learn more about the subject by reading books about autism and life stories of those who live with it.

The best example I found were in the books by Mary Temple Grandin. Temple, a scientist in the field of animal behavior, describes her journey through life as an autist in her book “Thinking in Pictures”, that eventually became an HBO movie. Ms. Grandin has been largely responsible for clarifying to the public what is to be autistic and the challenges she has faced in her upbringing and career. She published a number of books on the subject, including explorations of the differences of the autistic brain and practical advice for parents and teachers (GRANDIN, 2008; GRANDIN, 2011; GRANDIN; PANEK, 2013).

Eventually I stumbled across psychological studies. Simon Baron-Cohen, a British psychologist, proposed this idea of “mindreading”, a theorized capacity of the mind based on a model named *Theory of Mind* (ToM), and the thesis that people with autism did have some sort of deficiencies in its development. For a parent, knowing that there was a theory and people had been studying it in academia for some time was, at the very least, conforting (BARON-COHEN, 1997).

After that first phase of getting to know more about autism, one fact became quite clear to me: Autism has no cure. Therefore, my attention was directed at finding ways to help. When I found myself coming back to academia, I thought to myself: How can I contribute to the research in artificial intelligence to bridge the gap with assistive systems for Autism? This is what motivates this work.

1 Introduction

1.1 The Autism Enigma

Autism Spectrum Disorder (ASD) (WHO, 1993) is a biologically based neurodevelopmental disorder characterized by marked and sustained impairment in social interaction, deviance in communication, and restricted or stereotyped patterns of behaviors and interests (KLIN, 2006). ASD prevalence in Europe, Asia, and the United States range from 1 in 40 to 1 in 500, according to population and methodology used (AUGUSTYN, 2019). The prevalence of autism in Brazil is estimated at 500,000 individuals (0.29%, according to the 2000 census). However, data is sparse and hard to come by due to the lack of consistency in recent epidemiological studies (MARTINS et al., 2014; PAULA et al., 2011; VALADÃO et al., 2016).

The cost for families to raise and support an autistic child without an intellectual disability is high: around 1.4 million US dollars in the United States or the United Kingdom. Costs for special education services, parental productivity loss, and adult residential care all factor into this (BUESCHER et al., 2014; LAVELLE et al., 2014). In Brazil, data is again sparse, but anecdotal evidence suggests that costs can run as high as two times as much as a neurotypical child since there is little in public services to the general population.

Autism is a spectrum; that is, there is a range of higher and lower functioning disorders. Individuals in the spectrum face lifelong challenges with social interaction and communication. Even for the high-functioning individuals in this spectrum, life can still be challenging due to deficits in comprehending social cues and interaction. As Frith (2003) aptly puts it in her book “Autism: Explaining the enigma”, there is a lot we do not know yet. Research is active into the neurodevelopmental causes, the genetic makeup, psychological deficits, and many others.

1.2 Motivation

In the article “Curing Robot Autism: A Challenge”, Kaminka (2013) states that “Almost all robots are autistic; very few humans are”, in the sense that robots are pretty deficient in handling social situations. The article intends to challenge the AI research community to create the building blocks required for social intelligence. There is one remarkable coincidence here: research on autism and artificial intelligence had their early starts around the same time, around the 1940s-1950s.

Computer Science and, more specifically, AI research has approached autism in several ways, including interactive environments, virtual environments, avatars, robotics, and

technologies to assist with emotion recognition and understanding. Historically, Papert (1990) started exploring computational systems applications in education during the 1960s through the development of the LOGO language with the guided exploration of the environment by a digital "turtle".

From these earlier achievements, interactive environments, particularly computer games, are used as controllable environments to encourage interaction between autistic children, finding applications in therapy and as educational aids. Following this trend, virtual environments are used as social training tools by creating 3D simulations of typical interactions a child is likely to experience. In virtual spaces, avatars with a capacity for emotional expression are used to improve the social skills of autistic children (BOUCENNA et al., 2014; JALIAAWALA; KHAN, 2020).

Robots as a physical personification of avatars are employed as therapeutic tools to allow learning by imitation, recognizing movements, and helping autistic children recognize facial emotion. Applied Behavior Analysis (ABA) therapy, a widely used technique in psychology, can be used through Socially Assistive Robots with significant results. It is now understood that autistic children feel more comfortable interacting with robots due to the predictability of their actions (DAUTENHAHN; WERRY, 2004; DICKSTEIN-FISCHER et al., 2018).

In the area of emotion recognition and understanding, a recent review by Lima et al. (2019) outlines the efforts for the creation of software systems. The lack of identification of emotions is one of the critical deficits that can be readily observed in some autistic individuals. This is usually associated with the difficulty of keeping eye contact and is believed to bring about problems with communication and social understanding and interaction.

However, even with all the advances we have had up to this point, computational systems in the form of avatars or robots cannot understand humans and anticipate their intentions. Without these essential enablers, AI systems will not be able to assist humans - they will be little more than caretakers with a fixed timetable. We lack computational assistive systems for helping people, *in real time*, in the autism spectrum with their impairments in social interactions. In pursuing the development of such systems, we will address fundamental unresolved problems in artificial intelligence.

1.3 Our Research Question

The research question that drives the present work is: What would it take to implement theory of mind in a computational system? It is a tall order, sure, but such a system would allow us to create *Social Observers* capable of assisting people in the autism spectrum on how to properly analyze the environment, interpret social cues and provide assistance. Of course, expert systems that implement these cognitive systems would not be limited to assistants - we would be able to equip robotic systems with a critical requirement for human-robot interaction.

1.4 Our Vision

We propose the design of systems with the capacity to analyze environmental and visual social cues that the individuals in the spectrum do not readily interpret. Our long-term goal is for these systems to become able to provide expert advice on the best alternative for interaction to improve the outcomes of social integration for these individuals. The design of such systems could be based on the existing mechanisms that the human mind uses for facilitating social interaction.

It became clear to us that in order to achieve that, we need to understand human cognition. Autism research offers us an interesting approach: Theory of Mind (ToM). ToM is an internal cognitive system in our minds that allows us to understand other people by assigning mental states to them. Psychology researchers have, since the 1990s, proposed that one of the causes for deficits observed in individuals in the autism spectrum is likely caused by deficiencies in the makeup of this cognitive system in our minds.

1.5 Our Proposal

Our proposal is to create a first iteration, as a stepping stone, for what we think the expert systems we describe above will be based on. This initial implementation will be based on a psychological model of Theory of Mind and will consist of a cognitive architecture to create a computational representation of the human mind apparatus on what concerns ToM.

Our intention is to create an *Observer* that is able to analyze scenarios, identify people and objects in this environment and assign, through the evaluation of intentionality and object properties, a set of representations of mind states related to the scenario to each person.

1.6 Methodology

We started by surveying psychological models of the mind that we could implement to imitate the human cognitive apparatus regarding autism deficits. We selected the mindreading model from the literature as proposed by the British psychologist Simon Baron-Cohen. An overview of the concepts we have explored can be seen in Chapter 2.

We then implemented an ad-hoc representation of the mindreading model. We describe this first effort, as well as the restrictions we found with it on Grassiotto and Costa (2021). After this initial effort, we understood the computational basis for the architecture. We decided to re-implement it using a more flexible cognitive architecture using the CST Toolkit as described in Chapter 3 in order to create a usable implementation that could be made available to the research community.

In order to validate this computational system, we defined test scenarios. We used

a representation of the canonical false-belief task, a test commonly used to identify autism spectrum deficits, (BARON-COHEN et al., 1985) and a set of proxy tasks that evaluate reading comprehension via question answering proposed by Meta Research (WESTON et al., 2016).

Due to the lack of availability of datasets with false-belief tasks in the literature, there were not also readily available scoring systems for the performance of computational systems in evaluating such scenarios. Because of that, we decided to create a new set of evaluation criteria, as outlined in Chapter 4 for the outcomes of the tests we ran and ranked the results according to these criteria. The architecture we defined was capable of passing the canonical false-belief task as defined by researchers in the autism spectrum. Additionally, it was capable of providing good results with a set of tasks usually used to qualify the capabilities of an artificial intelligence system, the Meta bAbI dataset.

1.7 Contributions

The main contributions of this work are:

- The creation of a computational-equivalent model of Theory of Mind. We have managed to take the basis of a theoretical model of the mind apparatus and implement a computational system inspired by its biological counterpart.
- The implementation of a cognitive architecture featuring the psychological model we described above. We created the necessary structures to express the computational model we defined using the CST Toolkit.
- The proposal of a scoring system for evaluating mental constructs created by the cognitive architecture. We described a set of rules to evaluate the outcomes of the computational system that would allow us to rank the effectiveness of the architecture we defined.
- The publication of the open-source code for the implementation of the cognitive architecture using the CST Toolkit (GRASSIOTTO; COSTA, 2020).

1.8 Publications

The present work resulted in the following publications:

1. GRASSIOTTO, F.; COSTA, P. D. P. Cogtom: A cognitive architecture implementation of the theory of mind. In: *ICAART (2)*. [S.l.: s.n.], 2021. p. 546–553.
2. GRASSIOTTO, F.; COLOMBINI, E.; SIMÕES, A.; GUDWIN, R. and COSTA, P. D. P. Cogtom-cst: An implementation of the theory of mind for the cognitive systems toolkit. In:

Proceedings of the 14th International Conference on Agents and Artificial Intelligence - Volume 3: ICAART,. [S.l.]: SciTePress, 2022. p. 462–469.

1.9 Organization

This work is organized as follows.

Chapter 2

In this chapter, we introduce to the reader some concepts central to our proposal.

Chapter 3

In this chapter, we will describe the architecture of the computational system we designed using the CST Toolkit, a rather flexible toolkit for creating cognitive architectures.

Chapter 4

In this chapter, our purpose will be to present the results obtained with the cognitive architecture we proposed in Chapter 3. We will first present the set of test tasks chosen for testing, then present the results we obtained.

Chapter 5

In this chapter, our objective is to outline our conclusions about the work and describe areas we think could be explored further.

2 Concepts

The objective of this chapter is to introduce to the reader some concepts central to our proposal. We start by describing autism, a neurodevelopmental disorder, the historical development of the modularity of the mind, the concepts of theory of mind, false-belief tasks, and mindreading. After that, we describe previous work on cognitive architectures, affordances, and the mechanisms for the understanding of intentions. We finish by exploring prior definitions of observers and their usefulness in artificial intelligence.

2.1 Autism

Autism was described independently and separately by Kanner et al. (1943) at John Hopkins University and Asperger and Frith (1991) at the University of Vienna in the 1940s as a condition affecting children, bringing forth unusual behaviors and deficits in social relatedness. We understand since the 1960s that autism is caused by a brain disorder present since infancy and found in all countries, social, economic, ethnic, and racial groups.

ASD, as pointed by the acronym, is a spectrum of disorders; that is, individuals in this spectrum can feature higher or lower functioning manifestations of the disorder. Asperger's Syndrome, a high-functioning form, is mainly characterized by impairments in social interactions and restricted interests and behaviors. Its development course is marked by the lack of clinically significant delays in cognitive development self-help skills and is usually associated with the development of typical intelligence. Lower-functioning or severe autism, on the other hand, frequently includes severe deficits in communication skills and lack of response to social interactions (KLIN, 2006; LORD et al., 2000).

Approaches to autism research include screening procedures, diagnostics, interventions, health issues, genetics, the biology of autism, and psychology.

Research in psychology includes cognition, perception, clinical research, neuroscience, and social psychology. One particular approach that finds applications to cognitive science and computer science is to explain the deficits observed in autism by identifying differences in the theoretical modular structure of the mind (FODOR, 1983).

2.2 Mind Modularity and Theory of Mind

The cognitive revolution of the 1960s led researchers to look at the mind in a new light. Instead of thinking of it as a black box, we started to think about the parts or processes that make it up. That was the birth of Cognitive Science, with the intent of joining the disciplines

of philosophy, psychology, linguistics, computer science, anthropology, and neuroscience. The basic tenet of Cognitive Science at the time was that *Cognition is Computation*. Computational approaches led to the concept of modularity, with the mental experience being explained as the result of multiple distinct processes rather than from a single, centralized one (BARRETT; KURZBAN, 2006; MILLER, 2003).

By that same time, Chomsky's studies in language acquisition had proposed the existence of a faculty of language, also known as a "language module", in the mind (CHOMSKY, 2011; CHOMSKY, 2017). In his view, the development of human language relied on the evolution of a unique computational system to process symbolic representations (HAUSER; CHOMSKY; FITCH, 2002).

In the 80s Fodor (1983), inspired by neo-Cartesianism and Chomsky's research in language development, wrote the seminal book *The Modularity of Mind*. In his work that defined the research in modularity for the following decades, Fodor maintained that only peripheral sensory systems were modular. In contrast, high-level cognitive functions were not (contrary to Chomsky's view). Today cognitive scientists defend the thesis that the mind is massively modular, and this thesis lead us to the psychology approaches to research into the Theory of Mind (CARRUTHERS, 2003; GALLISTEL, 2011; PALECEK, 2017).

Theory of Mind is also known in psychology research as *Commonsense Psychology* or *Folk Psychology*, and is concerned with the capability of explaining or predicting mental states of other people. There are four main approaches that we will briefly outline here: Theory-Theory, Modularity-Nativist, Rationality-Teleology and Simulation Theory (GOLDMAN, 2012).

The Theory-Theory approach was one of the earliest proposals that started in the 1950s with the notion that individuals have common-sense concepts and that children are capable, from an early age, to generate, test, and change theories about the physical and social world. The Modularity-Nativist approach, as the name suggests, proposes that one or more domain-specific modules exist in the mind that employ proprietary representations for the mental domain. The Rationality-Teleology approach tells us that one mind creates predictions or explanations by employing a system of norms, or rules for understanding the world. That is, we predict how one will believe or desire by assigning rules to what we should believe or desire. Finally, the Simulation Theory approach, also described as the *Empathy Theory* assumes that we can predict the behavior of other people by visualizing ourselves in their place. One common description is that we should answer to the question "What would I do in that situation?" to predict the mental states of others.

Following this brief outline, we come back to how Autism is related to the psychology models. In one of these approaches, Modularity-Nativist psychologists were concerned with explaining autism deficits. Theory of Mind (ToM), the innate human capacity of attributing mental states to others, was considered a crucial mechanism to enable human communication. Deficits in this mechanism were considered as a causative factor for autism (FLAVELL, 2004;

LAVELLE, 2012; PREMACK; WOODRUFF, 1978).

ToM research in this area started in the 1980s by analyzing children's performance in a set of tasks to measure the understanding of a peer child's beliefs as being false, known as False-Belief tasks. These are tasks used to check if the child understands that another person does not possess the same knowledge as herself.

In particular, Baron-Cohen et al. (1985) proposed the *Sally-Anne test* as a mechanism to infer the ability of autistic and non-autistic children to attribute mental states to other people regardless of the IQ level of the children being tested.

In the test, a sequence of images (Figure 2.1) is presented to the children. Starting the sequence, in the top rectangle, two girls (Sally and Anne) are in a room, with a basket (Sally's) and a box (Anne's). Sally takes a ball and hides in her basket (second rectangle), then leaves the room (third rectangle). After that, Anne takes the ball from Sally's basket and stores it in her box (fourth rectangle). Sally then returns to the room (fifth rectangle). The child is then asked, "Where will Sally look for her ball?" Most autistic children answer that Sally would look for the ball in the box, whereas control subjects correctly answer that Sally would look for the ball in the basket.

The results presented in the article supported the hypothesis that autistic children, in general, fail to employ a ToM due to the inability to represent mental states. The downside of this is that autistic subjects are unable to impute beliefs to others, bringing a disadvantage to predicting the behavior of other people. It is thought that this lack of predicting ability causes deficits in the social skills in people on the autism spectrum, making it much harder to face the challenges of social interaction.

Examples of abilities that are linked to the capacity of understanding each other mental states are, among others, the ability to empathize and the skills of coordination and cooperation (SALLY; HILL, 2006; SCHAAFSMA et al., 2015). ToM allows us to generate expectations about the behavior of others and, based on these expectations, guide our decision-making process.

As a first approach to the problem of interpreting environment social cues, this work embraced the challenge of implementing a computational model to act as an *Observer* of scenarios and to emulate theory of mind abilities. To reach this goal, we thoroughly analyzed the autism and ToM literature, and we identified that that the work of the psychologist Baron-Cohen (1997) and its mind-blindness theory of autism provides a modular framework that we found suitable for computational modeling. The following section describes the main aspects of the Baron-Cohen's Mindreading model.

Figure 2.1 – Drawing depicting a sequence of images for the Sally-Anne test for false-belief, a mechanism to identify theory of mind capabilities in children. Adapted from (BARON-COHEN et al., 1985), drawing by Alice Grassiotto.

2.3 Baron-Cohen's Mindreading

The British psychologist Baron-Cohen (1997) proposed the mindreading model in his book *Mindblindness-an essay on autism and theory of mind* as a modular, innate, and evolution-driven system that allows us to make sense of the actions of others. He proposed that the cognitive delays associated with autism are related to deficits in developing such a system. In the text, he suggests four mechanisms that would be required for the human capacity to mindread as can be seen in Figure 2.2.

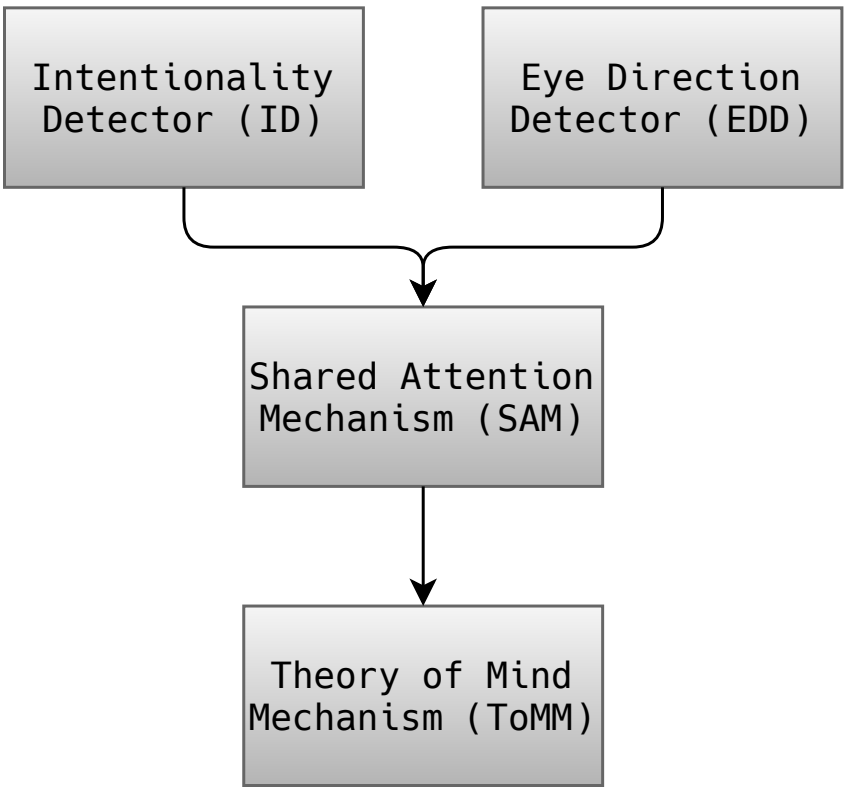

Figure 2.2 – The Mindreading Model as proposed by Baron-Cohen, showing the four mechanisms present in the human mind. Extracted from (BARON-COHEN, 1997)

- **Intentionality Detector (ID)** is a perceptual device that can interpret movement and identify agents from objects and assign goals and desires.
- **Eye Direction Detector (EDD)** is a visual system that can detect the presence of eyes or eye-like stimuli in others, to compute whether eyes are directed to the self or towards something else and infer that if the eyes are directed towards something, that the agent to whom the eyes belong to is seeing that something.
- **Shared Attention Mechanism (SAM)** builds internal representations that specify relationships between an agent, the self, and a third object. By constructing such representations, SAM can verify that an agent and the self pay attention to the same object.

- **Theory of Mind Mechanism (ToMM)** completes the agent development of mindreading by representing the agent's mental states that include, among others, the states of pretending, thinking, knowing, believing, imagining, guessing, and deceiving.

The components of the mindreading model are not isolated as there are interactions required to build the internal representations of the EDD, SAM, and ToMM.

The modular structure proposed by Baron-Cohen and presented here is concerned with high-level processing and creation of mind structures. For this work, it needs to be fit to a more comprehensible agent model and simulation system. This is where the concept of cognitive architecture becomes necessary.

2.4 Cognitive Architectures

According to Antonio Lieto et al., "Cognitive Architectures are both abstract models of cognition, and the software instantiations of such models, which are then employed in the field of Artificial Intelligence (AI)" (LIETO et al., 2018).

A recent review by Kotseruba and Tsotsos (2018) states that over the last 40 years, hundreds of Cognitive Architectures have been proposed. The research goal in the area is to model the human mind towards achieving human-level intelligence. Goertzel et al. (2010) states that BICAs, Biologically Inspired Cognitive Architectures, i.e., cognitive architectures that draw inspiration from the human brain, are distinct from those inspired by models of the mind, even though this distinction has become blurry over the years.

Cognitive Architectures seek to implement human-level intellectual abilities. Among these abilities, metacognition, described by the psychologist Flavell (1979) as "thinking about thinking", is the set of abilities that monitors internal processes and allows for reasoning about them. Metacognition is considered an essential requirement for social cognition and achieving ToM in a cognitive architecture.

Theory of mind mechanisms have not been the focus for most cognitive architectures we surveyed. We believe this is because these are high-level constructs that demand the comprehension of advanced cognitive processes that may require complex computational implementations. We review here the previous achievements of the research in the area.

The Sigma cognitive architecture, described by Rosenbloom (2013), Rosenbloom, Demski and Ustun (2016), seeks to implement an integrated computational model of intelligent behavior. Sigma has demonstrated an application for simultaneous-move games, particularly with the resolution of the Prisoner's Dilemma by employing combinatorial search (PYNADATH et al., 2013). Polyscheme explored perspective-taking for robots interaction with humans, but did not try to model ToM (CASSIMATIS, 2001; TRAFTON et al., 2005).

Anderson, Matessa and Lebiere (1997), Anderson and Lebiere (2014) described ACT-R theory (ACT: Atomic Components of Thought; R - Rational Analysis) as a means to achieve the unified theory of human cognition. ACT-R defines, through the concepts of *Chunks*, declarative knowledge, and *Productions*, procedural knowledge, as the minimal atomic parts of the human cognition.

The short report by Triona, Masnick and Morris (2002) defines in ACT-R 4.0 (an updated version of the architecture) a minimal model or the processes required for simulating the performance on false-belief tasks. This model defined as *Chunks* the goals for the information in a false-belief question, the general knowledge for knowledge relevant to the current question, and objects to identify object-specific information, and five *Productions*: two designed to respond to control questions, two that respond to the false-belief question and one to stop the model. The objective of this exercise was to use computation modeling, through the ACT-R cognitive architecture, to refine the understanding of children 's development of theory of mind into a testable environment. Their purpose was quite different from ours, in the sense that our target is rather the development of a machine equivalent of the theory of mind.

Scassellati (2001) proposed a novel architecture called "Embodied Theory of Mind". Scasselatti presented the theories of the psychologists Leslie and Baron-Cohen on the development of ToM in children, discussing the potential application of both in robotics (LESLIE, 1994; SCASSELLATI, 2002). His work had the initial objective of applying the psychological models to the sensory detection of human faces and identifying agents based on animated stimuli but did not proceed to implement higher-level constructs in the form of mind beliefs.

We conclude that from the cognitive architectures surveyed, Sigma, Polyscheme, ACT-R, and the "Embodied Theory of Mind" proposed integrating principles of ToM, in order to enable specific robotic behavior, simulating human-like social and interaction capabilities. However, we find that there is an opportunity to expand on the achievements of previous research.

The novelty of this work is to propose a unique cognitive architecture that seeks to integrate ToM mechanisms, inspired by the research in psychology, to enable the creation of a *Observer-like implementation*. Our observer has the long-term purpose of becoming an assistant to comprehend social cues for individuals in the autism spectrum.

2.5 The Cognitive Systems Toolkit (CST)

The Cognitive Systems Toolkit (CST) is a general toolkit for constructing cognitive architectures. There are clear advantages of selecting the CST for this work; due to its generic nature, it offers a flexible base that allows the implementation of problem-specific cognitive architectures as we are trying to achieve here (PARAENSE et al., 2016).

Inspired by Baars and Franklin's Global Workspace Theory (GWT) for conscious-

ness, Clarion, and LIDA cognitive architectures, among others, CST uses many concepts introduced there(BAARS; FRANKLIN, 2007; BAARS; FRANKLIN, 2009). GWT establishes that human cognition is achieved through a series of small special-purpose processors of an unconscious nature. Processing "Coalitions" (i.e., alliances of processors) enter the competition for access to a limited capacity global workspace (SUN, 2006).

The core concepts in the CST Core architecture are *Codelets* and *Memory Objects* as can be seen in Figure 2.3.

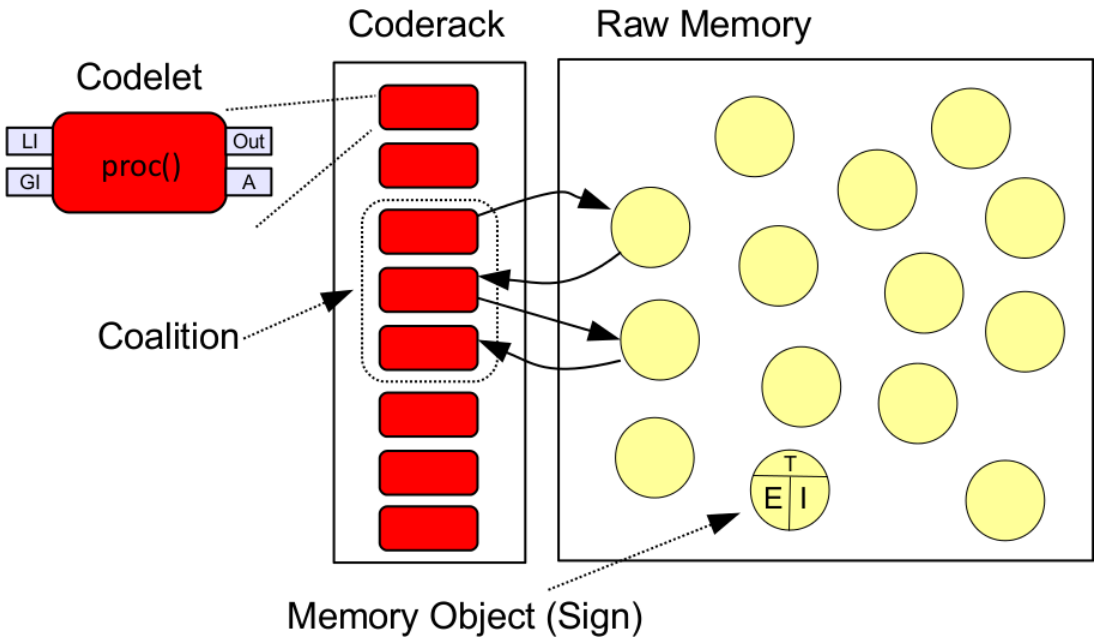

Figure 2.3 – The CST Core Architecture, consisting of codelets in a coderack and memory objects in raw memory. From https://cst.fee.unicamp.br/

Codelets are defined as micro-agents, small pieces of non-blocking code with a specialized function, designed to be executed continuously and cyclically for the implementation of cognitive functions in the agent mind. Codelets are stored in a container known as the *Coderack*.

Memory Objects are generic information holders for the storage of any auxiliary or episodic information required by the cognitive architecture. Memory Objects can also be organized in *Memory Containers* for grouping purposes.

In the CST Core, there is a strong coupling between Codelets and Memory Objects. Memory Objects are holders for any information required for the Codelet to run and receivers for the data output by the Codelet. In a similar fashion to Codelets, all Memory Objects and Memory Containers are stored in a container known as the *Raw Memory*.

2.6 Human Intention Understanding

In social cognition, researchers believe that an artificial intelligence's capacity to understand a person's intentions is necessary for human-machine interaction. By understanding environmental cues, an AI could deduce the human intention by considering the relationship between objects and actions. One example of understanding intentions is to detect if Sally, in the Sally-Anne test described earlier, intends to put her ball inside her basket.

There are some approaches to achieve this: visual action classification through recursive neural network models, object affordance-based intention recognition, and inference through bayesian models of hierarchical plans (YU et al., 2015; HOLTZEN et al., 2016). The first concept worth exploring for intention understanding is affordances.

Behavioral psychology and artificial intelligence have had a long history of cross-pollination between them. One insight from psychology, the concept of affordances and their perception in human behavior, inspired AI practitioners to extend it. The psychologist Gibson (2014) introduced the concept of Affordances in his 1979 book *The Ecological Approach to Visual Perception*. In his words (including italics):

> "The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill. The verb *to afford* is found in the dictionary, but the noun *affordance* is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment."

The importance of the concept of affordances - a term Gibson himself created - cannot be stressed enough. It establishes that the way the living perceive the world is directly related to the actions that can be performed in the environment.

In computer science, the formalization of affordances was the object of research in AI and robotics. Steedman, a computational linguist, sought to create a formal theory, using linear dynamic event calculus, for the relationship between events in the environment the object-actions pairs, applicable to human cognition. Şahin et al. (2007) reviewed classic formalization proposals (Turvey's, Stoffregen's, and Chemero's) and introduced a new one with the purpose of application to autonomous robot control. In their proposal, the concept of the effect to the environment of an action of an agent on an entity was explored (TURVEY, 1992; STEEDMAN, 2002; STOFFREGEN, 2003).

Affordances are a powerful concept. Researchers have found applications for affordances in robotics as a process to encode the relationships between actions, objects, and effects. Classic examples are that a ball might afford to catch, or a box might hide something inside (MCCLELLAND, 2017; MONTESANO et al., 2008). They are, however, only part of the picture for an AI that targets understanding the environment. If we again look at the human

model as inspiration, the theory of mind relies on lower-level capacities of the brain, such as the capacity of understanding other people's intentions through the observation of actions. This innate capacity, of interest to this work, is known as *Intention Understanding* (BLAKEMORE; DECETY, 2001).

Intention understanding is a core part of the mindreading process. We know that humans are very good at identifying intention; by 18 months of age, babies can associate intentions with observed actions. In AI, two approaches are used, namely, plan recognition and visual activity recognition through matching to intent libraries. Plan recognition, the problem of determining an agent intention through mapping of the agent action, has primarily used probabilistic distributions to identify "explanations" for a set of actions, whereas visual human activity recognition (HAR), a field of computer vision, uses hierarchical frameworks starting from pixel-based detection to high-level reasoning engines (BONCHEK-DOKOW; KAMINKA, 2014; CHARNIAK; GOLDMAN, 1993; SUBETHA; CHITRAKALA, 2016).

2.7 Ideal Observer Theory

Ideal observer theorists in the field of Ethics characterize Ideal Observers as entities with characteristics to eliminate incorrect reactions of humans due to the lack of information, undue bias, or inconsistencies in general. Ideal Observers would ideally be well informed, impartial, consistent, and empathetic (KAWALL, 2013).

The roots of Ideal observer theories come from 18th-century philosophy, but contemporary definitions by Firth (1952) have left a legacy today, even in the field of Artificial Intelligence. For Firth, anything that can be accepted as right or wrong is defined by the ideal observer's reaction to an act.

Inspired by Firth's theory, Savulescu and Maslen (2015), Giubilini and Savulescu (2018) discussed the concept of a *Moral Artificial Intelligence (MAI)* to advise human agents to select the right course of action through monitoring factors that affect moral decision making.

2.8 A Caveat: Embodied Cognition and Body Schemas

Embodied Cognition is a field of research that emphasizes the importance of the physicality of an agent's body in its cognitive abilities. Related to this, Body Schema is the wide-ranging term used in a number of fields, particularly robotics, for the supporting processes that map out the posture of the body in an environment.

According to Spaulding (2014) on the article "Embodied cognition and theory of mind":

> "Embodied cognition proponents reject the idea that social cognition is based on

> ascribing mental states to others. On their account, the capacity for more basic, non-mentalistic, interactive embodied practices underlies our ability to understand and interact with 1 others. These interactive embodied practices consist in primary intersubjectivity and secondary intersubjectivity."

That is, Embodied Cognition offers a counterpoint to the mindreading concept in the sense that it rejects the notion that social cognition is based on the ability of assigning mental states to other people. According to researchers in the area, the capacity of interactive embodied interaction is necessary for the ability of understanding other people (SHAPIRO, 2014).

We see Embodied Cognition as a theoretical approach that does not directly relates to the work we introduce here since our proposal is for an expert Observer that does not either employ movement in the environment or interact physically with the entities in a scenario.

2.9 Concluding Remarks

This chapter's objective was to discuss the concepts relevant to his work.

The main driver for this proposal is autism, a neurodevelopmental disorder, with unique challenges for social adaptation. We introduced concepts from cognitive science and psychology to explain the development of modularity theory and research into the theory of mind.

We discussed how cognitive architectures, abstract models of cognition, and their implementations, have approached the subject of simulating cognitive functions of the mind, including metacognition and the theory of mind. We described the Cognitive Systems Toolkit, a versatile toolkit for creating cognitive architectures.

We proceeded to talk about Intention Understanding and Affordances, innate capabilities of the human mind that are a requirement for artificial systems to understand the world. We finished by looking at the original concept of the Ideal Observer in Ethics and an application in moral advisors in AI.

Our proposal is, inspired by the research in cognitive science and psychology, to create a novel cognitive architecture that would simulate the human cognitive apparatus necessary for social interaction. We looked into the models for the theory of mind and integrated them with concepts already explored by the artificial intelligence and robotics community to achieve our objective. Our end goal is to create an *Observer* capable of assisting individuals in the autism spectrum on what regards observations of social interaction.

In the next chapter, we will describe the architecture of the computational system we designed.

3 Computational System Description

The present work focus on the computational proof-of-concept of the ToM psychological model proposed by Baron-Cohen (1997). In this chapter, we present and describe the main components of CogToM, a cognitive architecture built using the Cognitive Systems Toolkit (CST), a general toolkit for the construction of cognitive architectures.

Baron-Cohen's ToM theory relies highly on the processing of basic elements of information that are the output result of sophisticated processes of our brain's perceptual system. This is the case, for example, of the mechanism of attention, which enables the recognition of the relevant objects, people, actions, and episodes in a scene. According to Baron-Cohen (1991), attention, and joint attention, which occurs when two people share their attention towards the same focus of interest, are critical aspects of a fully-fledged theory of mind.

However, the present work does not implement an associated complete perceptual system. We assume that functionalities such as people and object detection, affordance recognition, and intentionality processing are implemented by third-party algorithms, such as those mentioned in Section 2.6. Section 3.4 describes how we emulate the existence of such sophisticated computer vision models through external inputs that provide the information expected from visual processing algorithms. As a toolkit, CST provides us with the tools for representing generic processing systems, so we can focus on creating an overall structure inspired by the psychological model, validating it, and then proceed to the implementation of the individual codelets in the architecture, without altering the base principles of the model.

We will start by describing our objective of implementing an *Observer*. Then we will explain how the CST toolkit could be used to achieve this objective. After that, we will explore the modules we implemented inspired by a psychological model, and we finish by describing the inputs and outputs of the system.

3.1 Architecture Description

Recent work by Savulescu and Maslen (2015), Giubilini and Savulescu (2018) discussed the concept of developing a *Moral Artificial Intelligence (MAI)* to advise human agents to select the right course of action through monitoring factors that affect moral decision making. We find strong parallels between the idea of MAIs and our proposal here.

The proposed architecture, as we implied, implements an AI *Observer* as well. The objective of this *Observer*, rather than serving as a moral compass (with all the limitations and restrictions that such a system would have), is to advise people on the autism spectrum of the social cues related to an analysis of the environment.

3.2 System Architecture

We have designed this cognitive architecture using the CST toolkit. Within this toolkit, one cognitive architecture can be modeled by defining Codelets and Memory Objects.

Codelets are the processes executed within one simulation step. Codelets have local (LI) and global (GI) inputs and provide an activation (A) and a set of outputs (O) to Memory Objects, as can be seen in Figure 3.1.

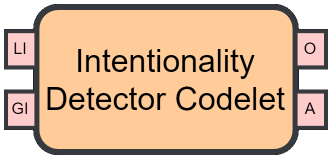

Figure 3.1 – An example codelet, the Intentionality Detector Codelet. Taken from our proposed architecture.

The second construct from the CST toolkit we use in this system is **Memory Objects** (and groupings of them, known as Memory Containers). These are generic information holders in memory that are modeled after the storage of data required for the execution of Codelets, as can be seen in Figure 3.2.

According to Baddeley, Eysenck and Anderson (2014), the classification of memory into a number of subtypes as we see in most cognitive architectures remain controversial. However, the standard multi-store model proposed by Atkinson and Shiffrin is widely used (ATKINSON; SHIFFRIN, 1968).

In this model, memory is divided into three main stores: sensory memory, short-term memory (STM) and long-term memory (LTM), and information is transferred between these stores in a linear fashion.

- **Sensory Memory** is memory for the brief storage of information associated to a single modality (auditory, tactile, etc). Sensory memory is usually more related to perception that to memory itself.
- **Short-Term Memory** is memory used for the retention of small amounts of information after brief delays. The concept of working memory is associated with STM, and is necessary for temporary maintenance and manipulation of information.
- **Long-Term Memory** is our main system for the storage of information for long periods of time. LTM can be explicit, or declarative, or implicit or non-declarative. Explicit memory is memory we can intentionally recall, and include semantic and episodic memory, whereas implicit memory is memory linked to learning processes.

  Semantic memory is usually related to factual and conceptual knowledge about the world and has an integration with the expression of language (QUILLAN, 1966). Episodic

memory is seen as the memory for one's life events, one's recollections and experiences (TULVING, 1986).

Our system employs sensory memory, in the form of external input systems, short-term memories in the form of working memories and percepts, and long-term memory as semantic memories for the knowledge previously acquired necessary for understanding the interaction between agents and objects.

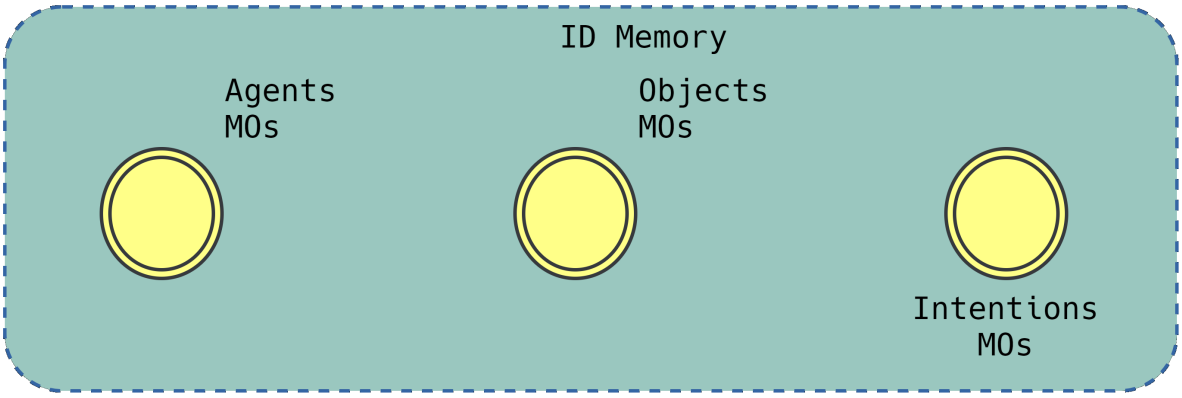

Figure 3.2 – A set of memory containers present in ID Memory - the Agents Memory Objects, Objects Memory Objects and Intentions Memory Objects. Taken from our proposed architecture.

We have designed this cognitive architecture for an agent that implements decision-making processes to implement an Artificial Intelligence *Observer*. The objective of this *Observer* is passing a false-belief task by implementing the mindreading model and integrating it with the processing of affordances and intentions.

The mindreading model, as described before, proposes four separate modules, or mechanisms, to implement the Theory of Mind functionality as can be seen in the Figure 2.2 presented before.

3.2.1 The Intentionality Detector Module

The starting point for our model is the Intentionality Detector module **(ID)**. The external inputs, the ID Codelet, and the Memory Objects for the ID module are represented in Figure 3.3.

As Baron-Cohen told us, he had no intention of making us mistake it with Freud's id concept from psychoanalysis. However, he hints that they both share the same pronunciation. Our ID is a perceptual device that, according to Baron-Cohen's psychological model, can interpret motion in the environment as it relates to the mental states of goal and desire. These mental states are primordial, as they are needed to identify approach and avoidance movements. These movements are essential in the sense that they are required for survival.

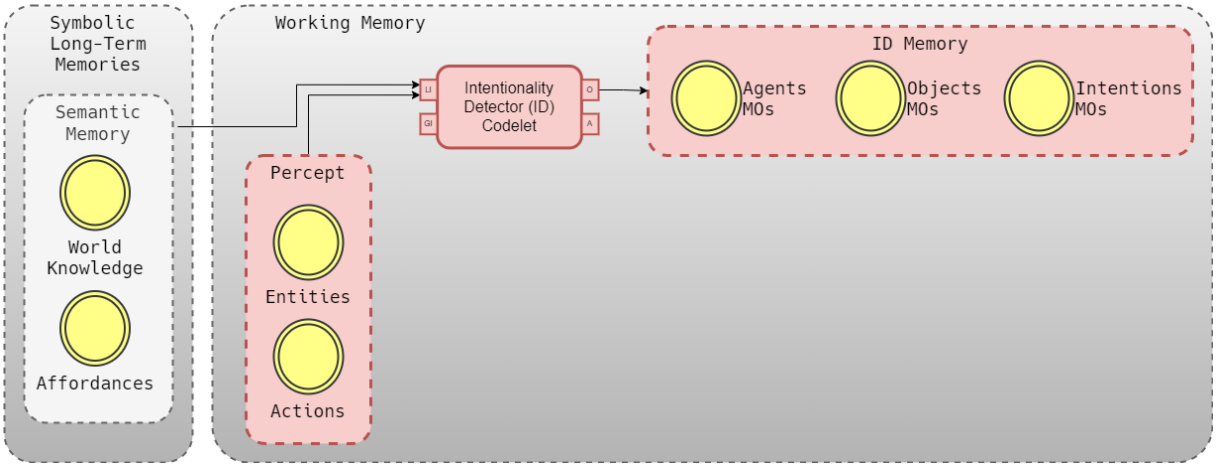

Figure 3.3 – The Intentionality Detector Module as a set of memory objects in working memory and the ID codelet. This is inspired by Baron-Cohen's psychological model.

Activation of this module is supposed to occur when there is, in the environment, any input that can be identified as an agent, that is, something that can move by itself. Some objects can undoubtedly be mobile, but ID can easily discard false agents by analyzing goals and desires.

ID is modeled as a Codelet in the system. **The Intentionality Detector Codelet** identifies which entities in a scene are agents or objects, based on movement and action detection, creating memories for the Agents, their Intentions, and Objects. An object here is defined as any entity in the scene that was not identified as an agent.

ID takes external inputs as memories that are not implemented by this architecture. Semantic memory provides prior knowledge about the world and affordances for common objects (World Knowledge and Affordances Memory Objects). In contrast, percepts feed ID with the perceived entities (agents or objects) and the actions being taken by these entities in the Entities and Actions Memory Objects. These are a model for the mental states of goal and desire described above.

The memory objects for the Actions percept is simplified in this implementation of the architecture. It is used here to convey only if an agent or object is active or not, by setting a boolean flag indicating, through the perception of drive-initiated action, if the entity in the entities input table is an agent or an object, as can be seen on Table 3.1. The column 'Is_Agent' informs the system that the entity has been perceived as an agent due to the actions perceived in the environment.

Semantic memory information is also simplified in this implementation of the architecture, consisting at this point of the affordances associated to each entity in the scene, as can be seen on the Table 3.2. The box entity is supposed to offer the containability property to the environment, whereas the ball can be hidden, and Anne (as an agent) has the basic property of being able to exist.

Table 3.1 – Example Entities Table

Entity	Is_Agent
Sally	True
Basket	False

Table 3.2 – Example Affordances Table

Entity	Affordance
Box	Contains
Ball	Hides
Anne	Exists

Memory Objects in **ID memory** are the agents, their intentions, and objects in the environment. Each of these entities is modeled as Memory Containers (MCs) in the architecture, producing an Agents MC, Objects MC, and Intentions MC.

The Agents MC consists of a grouping of Agent Memory Objects, where each memory object consists of a symbol of a single agent; the Objects MC consists of a grouping of Object Memory Objects, where each memory object consists of a symbol of a single object; finally, the Intentions MC consists of a grouping of Intention Memory Objects, with each intention being represented by four symbols: an agent, an intention, an object, and a target object for the intention.

3.2.2 The Eye-Direction Detector Module

The Eye Direction Detector (EDD) is the second mechanism found in the model. The inputs from ID memory, the EDD Codelet, and the Memory Objects for the EDD module are represented in Figure 3.4.

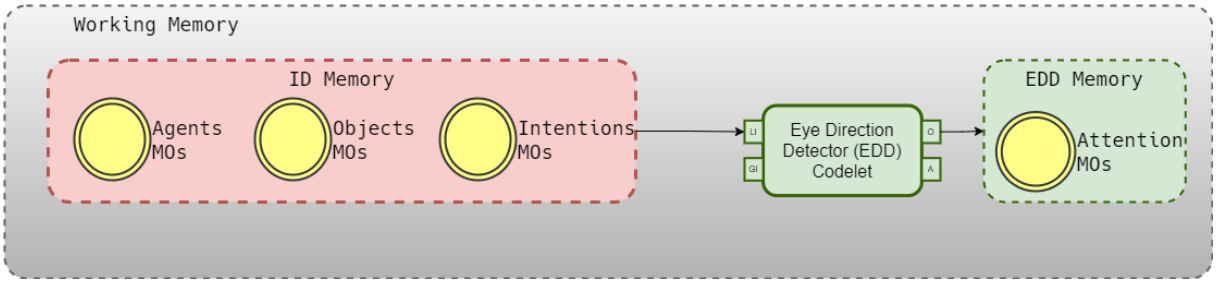

Figure 3.4 – The Eye-Direction Detector Module as a set of memory objects in working memory and the EDD codelet. This is inspired by Baron-Cohen's psychological model.

According to the theory, EDD is a specialized part of the human visual system, with

three core functionalities: the detection of the presence of eyes, the computation of whether said eyes are directed towards the self or some other entity in the environment, and if the entity in the environment is capable of identifying that the eyes are directed towards itself. This last capability is vital for children to attribute perceptual states to other people, as the classical "My mom sees me".

EDD is also modeled as a Codelet. **The Eye Direction Detector Codelet** identifies eye direction from the agents and objects created by the ID Codelet creates and attaches that information to Attention Memory Objects. This Codelet is activated by the ID Codelet.

Memory Objects in **EDD memory** store the attention of each agent for objects or other agents in the environment, modeled as Attention MCs.

Attention MCs are grouping of Attention Memory Objects, that are comprised of pairing of two symbols: a first symbol that represents the agent that is paying attention, and another symbol for the object or person that is the target of the attention of the first symbol.

3.2.3 The Shared Attention Mechanism Module

The Shared Attention Mechanism (SAM) is our third module. The inputs from EDD memory, the SAM Codelet, and the Memory Objects for the SAM module are represented in Figure 3.5.

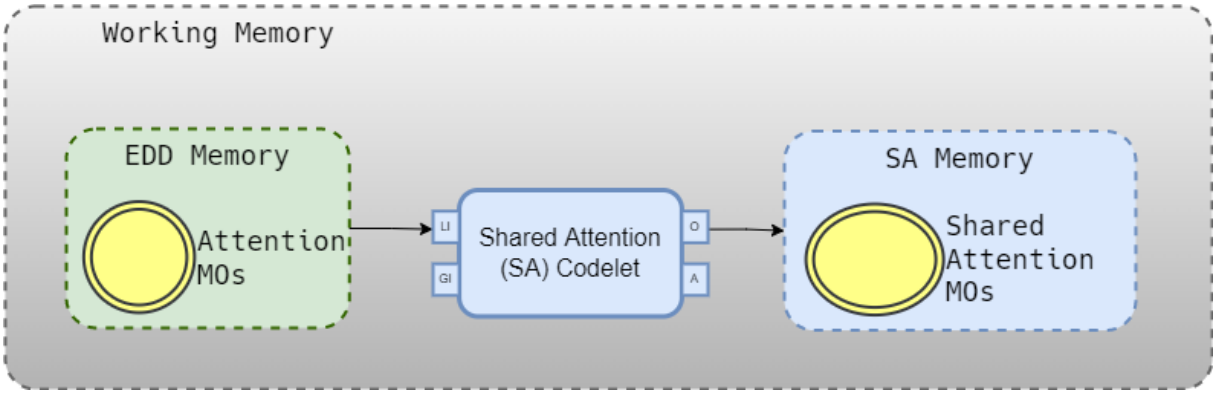

Figure 3.5 – The Shared Attention Mechanism Module as a set of memory objects in working memory and the SAM codelet. This is inspired by Baron-Cohen's psychological model.

SAM builds representations for the relationship between an agent, the self, and a third entity, building upon EDD information. These relationships are called **triadic representations** since they relate to a set of three elements. What do these representations encode? In practice, their purpose is to specify shared attention. In other words, the self and another agent are paying attention to the same object.

These triadic representations according to Baron-Cohen (1997) are of the form

$$< Agent/Self, Relation, Self/Agent - Relation - Proposition >$$

For example:

$$< Sally\,(Agent), sees\,(Relation), Anne-sees-the-ball\,(Agent-Relation-Proposition) >$$

SAM is a critical capability for us. One example of "SAM in action" is when you are playing a competitive sport; it is essential to identify that you and another player are paying attention to where the ball is.

SAM's Codelet, **The Shared Attention Mechanism Codelet** detects shared attention from the objects created by the EDD Codelet and then creates and attaches information to Shared Attention Memory Objects. This Codelet is activated by the EDD Codelet.

Memory Objects in **SAM memory** store which objects or agents have the shared attention of two or more agents, modeled as Shared Attention MCs.

Shared Attention MCs are grouping of Shared Attention Memory Objects, that are comprised of a list of symbols representing agents that are sharing the attention, and a single symbol representing the object or agent that has the attention for that list of agents.

3.2.4 The Theory of Mind Mechanism Module

Our model now needs one final mechanism to complete the child's mindreading system: the Theory of Mind Mechanism (ToMM). All the inputs from the system, the TOMM Codelet, and the Memory Objects for the TOMM module are represented in Figure 3.6.

ToMM is a system to fully represent *epistemic* (that is, related to knowledge or knowing something) mind states. These mental states include pretending, thinking, knowing, believing, and so on.

ToMM has to be able to create a *theory*, that is, a procedure to bind together all the inputs from the other modules of the model with the knowledge-based mental states. It is believed that ToMM creates representations of the form

$$< AGENT, ATTITUDE, PROPOSITION >$$

For example, "Sally BELIEVES Ball Hidden in Basket".

ToMM is implemented as a Codelet. **The Theory of Mind Codelet** works as an integrator of all the information from working memory and creates Belief Memory Objects. This Codelet is activated by all the previously described Codelets.

Memory Objects in **ToM Memory** are *Beliefs*, the main purpose of this cognitive architecture, modeled as Belief MCs.

Belief MCs are grouping of Belief Memory Objects. These are comprised of a list of symbols representing the agent and the object associated to this belief, a symbol for the

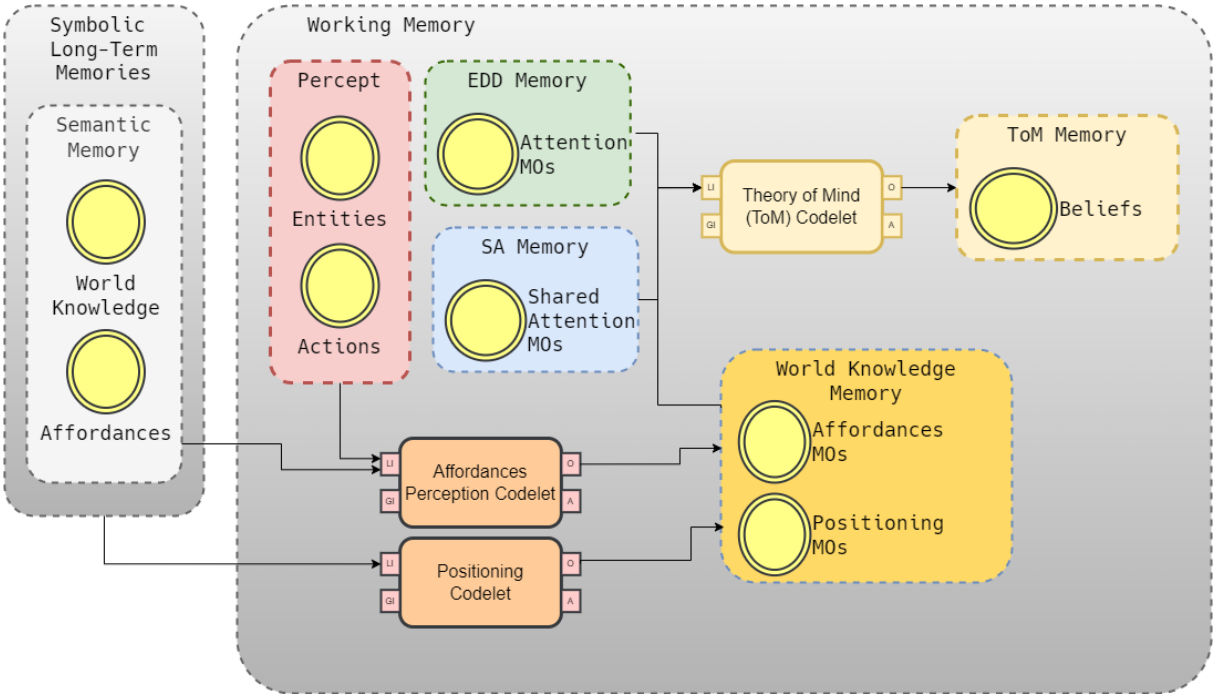

Figure 3.6 – The Theory of Mind Mechanism Module as a set of memory objects in working memory and the ToM codelet. This is inspired by Baron-Cohen's psychological model.

mental state that can either be *believing* or *knowing*, a symbol for the affordance related to the object, and a symbol for the target object related to this belief. Not all beliefs have a target object (example: Anne-Believes-Sally-Exists) and thus the target object symbol may be absent.

3.2.5 Auxiliary Subsystems

The **Affordances Perception Codelet** and the **Positioning Perception Codelet** do not exist in the mind model. Their purpose is to create affordances and position properties for the objects in the scene based on the camera input, the Agents, the Objects, and the Intentions associated with them.

Affordances Memory Objects are Memory Objects that retain agents and objects interaction properties as a dictionary lookup.

Positioning Memory Objects are Memory Objects created from a camera input to inform the current location of agents and objects in a scene.

Activation Memory Objects are special-purpose Memory Objects used in this architecture to synchronize the execution of Codelets. They are used to trigger the exact moment a Codelet should be executed, since the model requires a sequential behaviour, Activation memory objects are used to indicate that the conditions for executing a particular are due, and the codelet can be safely executed. They are not shown in the diagram above for simplification purposes;

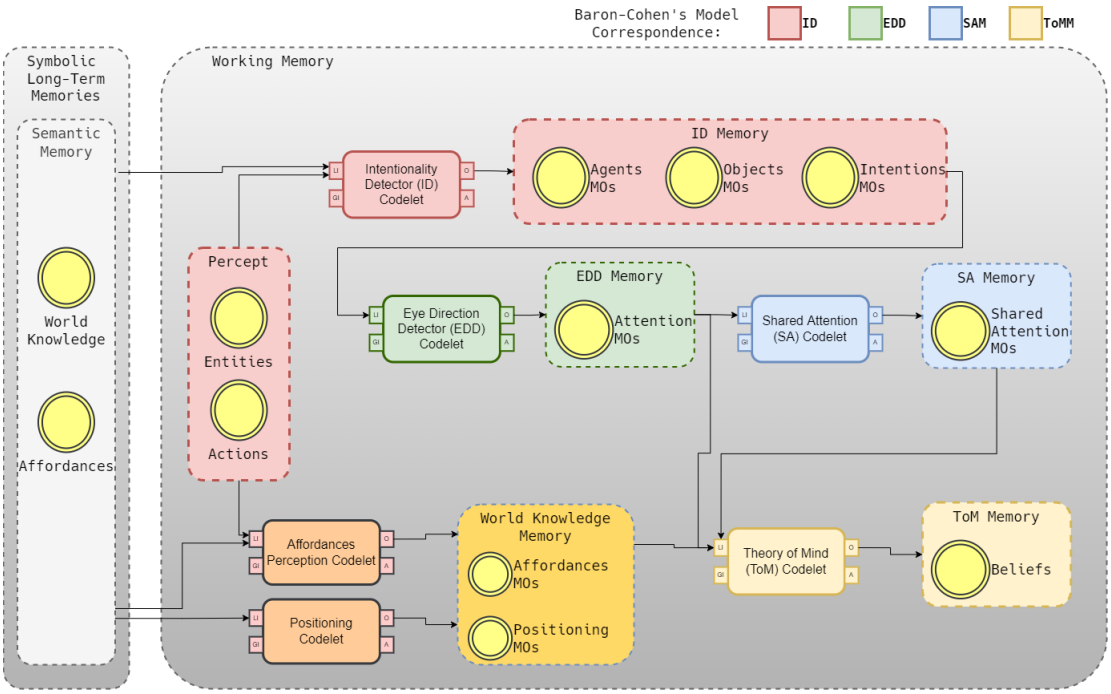

Figure 3.7 – The CogToM cognitive architecture, modeled on the CST Toolkit using codelets and memory objects.

3.2.6 The Big Picture

Putting it all together, the architecture we presented up to this point is modeled by defining Codelets and Memory Objects, as shown in Figure 3.7.

We designed our system to pass a false-belief task by implementing the mindreading model and integrating it with the processing of affordances and intentions. It relies on inputs as an external system, in the form of a visual camera system capable of identifying agents and objects, eye direction, and human intention and positioning. We see Affordances as properties of the entities (objects and agents) in the system.

The outputs of the system (*Beliefs*) are textual representations of the mental state of an agent as perceived by the *Observer*. The system is simulated by processing a set of inputs that are tied to temporal steps. These temporal steps are related to mind cycles and are defined as "Mind Steps".

3.3 Belief Construction

Beliefs in ToM Memory are modeled as descriptions for the mental states the Observer will provide. There are two sets of beliefs the system will consider: *Beliefs* for each one of the agents in the scene, and *self-beliefs*, those associated with knowledge the Observer has about the environment.

$$< AGENT, MENTAL STATE, OBJECT, AFFORDANCE, TARGET\ OBJECT >$$

Where:

- **AGENT** is the primary agent that the mental state applies to, for example, *Sally*. In the case of self-belief, the agent is the Observer itself.
- **MENTAL STATE** is the mental state assigned to the agent. Various mental states could be considered, including pretending, thinking, knowing, believing, imagining, guessing, and deceiving.

  For this implementation, we considered two mental states: believing and knowing. The mental state for believing, *Believes* is used for beliefs associated to the agents' beliefs, whereas the mental state for knowing, *Knows* is associated to the observer self-beliefs about the environment.
- **OBJECT** is the object of the belief, for example *Box*.
- **AFFORDANCE** is the main property, or affordance, of the object. For example, a Box may *Contain* something.
- **TARGET OBJECT** is the target object for the affordance, when applicable. For example, a Box may contain a *Ball*.

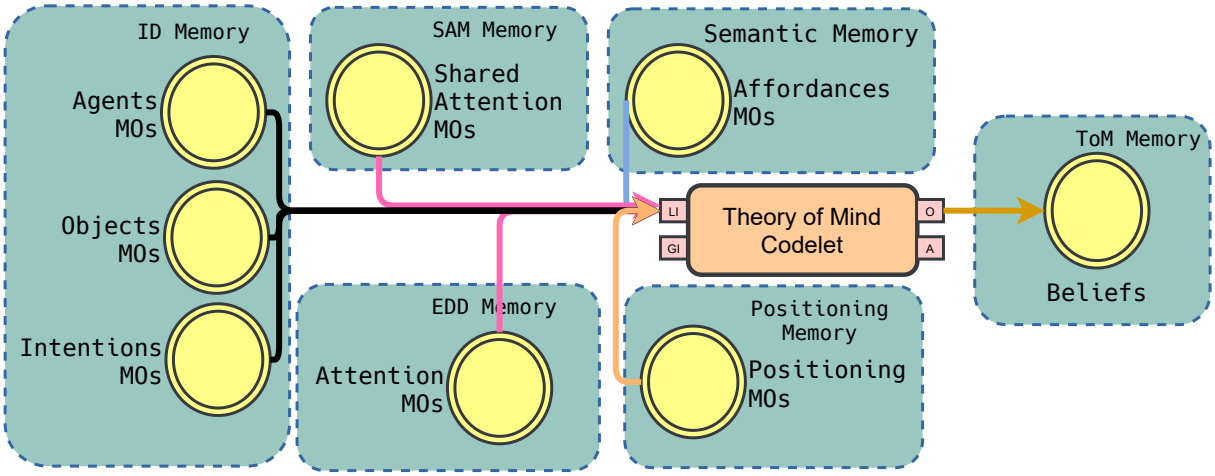

Figure 3.8 – A more detailed view on the interfaces to the ToM Codelet, showing interactions with the other modules (ID, EDD and SAM) memory structures for the construction of beliefs.

3.3.1 Rules and Inference Engines

In traditional artificial intelligence, inference engines are components of expert systems that apply reasoning to knowledge about the world to create new information. Our system did not create a generic inference engine to apply logic rules; instead, our approach here was to create a rules-based system as an approximation of the biological system we intended to emulate.

As shown in Figure 3.8, our rule-based system builds Beliefs from the ToM model proposed by the literature that provides the set of agents, objects, intentions, and attentions in the scene through the ID, EDD, and SAM modules and their Memory Containers. From this initial output, the architecture integrates the affordances from semantic memory. A single combination of an agent and one object defines a Belief object. Based on this set composed of an agent, an object, an affordance, and intention Memory Objects, a textual representation of that Belief is created in the ToM Memory as a new Memory Object. The set of rules we have hard-coded in this implementation creates associations by defining a pair of symbols, composed of the intention of an agent with the main affordance of an object, and assigns this pair of symbols with the final outcome of a belief.

Examples of the generated beliefs in ToM Memory are textual descriptions like the one on 3.3.1.

Listing 3.3.1: Example of a Belief

```
Sally BELIEVES Ball Hidden In Basket
Anne BELIEVES Basket Contains None
Observer KNOWS Sally IS AT Room
```

The first two beliefs, “Sally BELIEVES Ball Hidden in Basket” and “Anne BELIEVES Basket Contains None”, are associated with the agents in the scene. After processing the affordances and intentions, the Observer entity is associating with Sally and Anne’s beliefs related to the Ball and the Basket objects in the scene.

In the first belief case, the rules engine associates the intention of reaching for the basket with the property of the basket of providing a “hide spot” for the ball, and through this association the belief is built. Similarly, the second belief did not have an intention associated, so the belief the system creates contains only the basket affordance and the agent associated with it.

The third belief, “Observer KNOWS Sally is AT Room,” is the self-belief the Observer creates to identify the agents and objects in the scene. The rules engine here is simpler; it created a representation of static knowledge rather than the result of an association between one intention and one affordance.

Source code for the implementation of the CogToM Cognitive archictecture is open source(GRASSIOTTO; COSTA, 2020).

3.4 System Input and Output

The system we designed requires visual and dictionary inputs to simulate functionalities of the mind. We assume specialist systems would implement that. Visual inputs are defined as tabular data providing information about the environment. As an example, we will describe

the set of inputs we used to simulate two steps in the canonical false-belief test.

Scene descriptions are provided in human-readable format. Column *t* specifies the time-step of the simulation, whereas the *Scene* column describes how a human would characterize the scene.

t	Scene
1	Sally and Anne are in the room. Basket box and ball are on the floor.
2	Sally reaches for the ball.

Entities are described for each step in the simulation, qualifying what the agents and objects in a scene are. For each time step *t* this input describes a list of *Entities* and its classification as an agent or an object. This input aims to reproduce one of the functionalities of the ToM ID module, the capacity of telling apart objects from agents, based on intentional movement.

t	Entity	Is_Agent
1	Sally	True
1	Anne	True
1	Basket	False
1	Box	False
1	Ball	False
2	Sally	True
2	Anne	True
2	Basket	False
2	Box	False
2	Ball	False

Eye Direction input describes the human eye direction for each of the agents on a scene, identifying the entity the agent is currently looking at. For each time step *t* this input describes, for each *Agent* in the scene, what is the *Target* entity for the human gaze. Eye direction gaze has been researched extensively by psychologists as one of the pieces of evidence for autism and is implemented by the EDD module in the ToM model.

t	Agent	Target
1	Sally	Anne
1	Sally	Basket
1	Sally	Box
1	Sally	Ball
1	Anne	Sally
1	Anne	Basket
1	Anne	Box
1	Anne	Ball
2	Sally	Anne
2	Sally	Basket
2	Sally	Box
2	Sally	Ball
2	Anne	Sally
2	Anne	Ball
2	Anne	Basket
2	Anne	Box

Intention input describes an analysis of the probable intention of an agent on a scene. or each time step *t* this input describes for each *Agent* the most probable *Intention* to be acted in an *Object* with a given *Target*.

t	Agent	Intention	Object	Target
1	Sally	None	None	None
1	Anne	None	None	None
2	Sally	ReachFor	Ball	None
2	Anne	None	None	None

Positioning input provides the relative location data for each agent and object on a scene. For each time step *t* this input describes for each *Entity* the current *Location* in a scene. Positioning is important to allow an observer to identify the limits of attention for each agent.

t	Entity	Location
1	Sally	Room
1	Anne	Room
1	Basket	Room
1	Box	Room
1	Ball	Room
2	Sally	Room
2	Anne	Room
2	Basket	Room
2	Box	Room
2	Ball	Room

Affordances for an object or an agent in a scene are provided as dictionary inputs. This input describes for each *Entity* (objects and agents) in a scene the *Affordances*. Affordances, in this model, are immutable properties that do not change within the timeframe of the simulation.

Entity	Affordance
Box	Contains
Basket	Contains
Ball	Hides
Anne	Exists
Sally	Exists

We note that our approach for the affordances, of associating one entity to one affordance, may lead to a combinatory explosion, the classical frame problem described by McCarthy and Hayes (1981) when applied to first-order logic. This is an area for future improvement of the architecture.

The system outputs textual representations for the set of mental states associated with an *Observer* entity for the scene under analysis. For validation of the model, two sets of input data are described: one with the description of the entire set of interactions for the canonical false-belief task and a second set with descriptions of scene environments from the Meta bAbI dataset (META RESEARCH, 2021).

3.5 Concluding Remarks

In this chapter, we have presented the architecture for the cognitive architecture we are proposing. We described each one of the modules we designed based on the psychological model from Baron-Cohen. Then, we proceed to describe the inputs and outputs of the system.

In the next chapter, we will present the results obtained by the cognitive architecture we described in this chapter.

4 Evaluation and Results

In this chapter, our purpose will be to present the results obtained with the cognitive architecture we proposed in Chapter 3. We will first present the set of test tasks chosen for testing, then present the results we obtained.

4.1 Methodology

A vital aspect of the present work was defining a strategy to evaluate the cognitive architecture inspired by Baron-Cohen's mindreading model. We assumed the challenge: if the cognitive architecture models the ToM correctly, it should pass False-Belief tests. Our methodology of implementing a computational system capable of processing visual inputs and other dictionary-based databases and outputting textual representations of the environment under analysis.

Evaluating the success of a system in reasoning tasks can be approached using datasets of scenarios, as is the case in some research fields we highlight here. *Video Understanding*, for example, is primarily concerned with joint video and language understanding tasks with applications in video captioning and video question answering(GAN et al., 2017; HENDRICKS et al., 2017). *Social-IQ*, a benchmark designed to train and evaluate socially intelligent technologies, analyzes causal relationships in videos (ZADEH et al., 2019) and is of particular interest to this proposal.

Visual Question Answering (VQA) seeks to understand the high-level reasoning requirements for an AI to be able to reply to a natural language question about an image with a natural language answer (ANTOL et al., 2015). The objective of this work, in a similar vein, is to exercise the requirements for a computational system to be able to understand the visual environment and create a set of *beliefs* about it. However, the approach taken with VQA is quite distinct from ours because it seeks to employ a dataset of human-generated questions associated with images.

Physical and Mechanical Reasoning is research concerned with how people can understand a physical scene and replicate similar behavior computationally (BATTAGLIA; HAMRICK; TENENBAUM, 2013; LERER; GROSS; FERGUS, 2016). CLEVRER proposes reasoning tasks, such as description, explanation, prediction, and counterfactuals from a video source (YI et al., 2019).

In order to validate our proposal, we used a representation of the canonical false-belief task (BARON-COHEN et al., 1985). We searched on popular dataset search engines on the internet for the terms "false belief" (and some variations) for additional test sets. We obtained

the results on Table 4.1.

Table 4.1 – Dataset Search Results for False-Belief Tasks

Dataset Search Engine	Indexed Datasets	Number of Search Results
Google Dataset Search	31 million (2020)	about 100
Mendeley Data	29.3 million (2021)	35
Kaggle	50 thousand (2021)	0
Microsoft Research Open Data	100 (2021)	0

We could not find datasets with synthetic descriptions of other tasks of this nature in these results. The results we have found were not usable as synthetic descriptions of false-belief tasks; instead, they were primarily datasets of video files evaluating, in the fields of psychology and medicine, the results of standard testing using the canonical false-belief task as described by the literature.

We decided, then, to evaluate the CogToM architecture with a set of proxy tasks that evaluate reading comprehension via question answering proposed by Meta Research (WESTON et al., 2016). This set of tasks, organized in a dataset called bAbI, has the form of narrative episodes, or stories, together with questions about the state of the world and has become a benchmark dataset for reading comprehension.

The bAbI dataset comprises twenty tasks to evaluate distinct capabilities of artificial intelligence under test. The tests are classified according to the capability required for successfully replying to a question at the end of the test description. As an example, Meta bAbI task 1 is concerned with describing a *a Single Supporting Fact*, as can be seen below, and querying the system under test to reply correctly based on this single fact.

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary traveled to the office.
Where is Mary? A:office

4.2 Evaluation of the Proposed Architecture

We employed a set of test tasks to exercise the cognitive architecture under validation. These tasks describe scenarios where the observer entity analyzes the environment and creates mental constructs in textual beliefs. As we described in Chapter 3, we chose the canonical false-belief task and the Meta bAbI tasks for evaluating the system.

Meta Research bAbI project at Meta Research (2021) is organized towards the goal of automatic text understanding and reasoning. The dataset consists of 20 test tasks, with each task to test one skill that a reasoning system should have (WESTON et al., 2016).

We considered, then, these two sets of validation tests. First, the canonical false-belief task as described by the literature, and we picked a subset of tasks from the bAbI dataset from Meta Research.

4.2.1 The Sally-Anne test

As described before, Baron-Cohen et al. (1985) proposed the *Sally-Anne test* as a mechanism to infer the ability of autistic and non-autistic children to attribute mental states to other people regardless of the IQ level of the children being tested.

In the test, a sequence of images (Figure 2.1) is presented to the children. Starting the sequence, in the top rectangle, two girls (Sally and Anne) are in a room, with a basket (Sally's) and a box (Anne's). Sally takes a ball and hides in her basket (second rectangle), then leaves the room (third rectangle). Anne takes the ball from Sally's basket and stores it in her box (fourth rectangle). Sally then returns to the room (fifth rectangle). The child is then asked, "Where will Sally look for her ball?". Most autistic children, and neurotypical children under four years old, answer that Sally would look for the ball in the box, whereas control subjects correctly answer that Sally would look for the ball in the basket. Based upon this description, the sequence of steps in the test case below outlines the canonical false-belief test.

Canonical False-Belief Test

Sally and Anne are in the room. Basket, box, and ball are on the floor.
Sally reaches for the ball.
Sally puts the ball in the basket.
Sally exits the room.
Anne reaches for the basket.
Anne gets the ball from the basket.
Anne puts the ball in the box.
Anne exits the room, and Sally enters.
Sally searches for the ball in the room.
Where does Sally believe the Ball Is? A: Basket

4.2.2 Meta bAbI tasks

In "Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks," Weston et al. (2016) describe the first ten tasks for the evaluation of AI reading comprehension via question answering as we outline below:

Single Supporting Fact Task number one is a test to check if the AI system can identify, from a set of potentially irrelevant facts, one supporting fact that provides the answer for the question being asked.

Two or Three Supporting Facts Tasks two and three provide two or three supporting statements that need to be considered for the correct answer.

Two or Three Argument Relations Tasks four and five are designed to break standard NLP techniques, e.g., bag of words. The AI system needs to consider two or three relations in the text to answer correctly. The sentences describing the test case have reordered words in task four, whereas task 5 has statements in inverted grammatical order such as "Bill gave Jeff the milk."

Yes/No Questions Task six checks if the AI system is capable of answering correctly true/false questions.

Counting and Lists/Sets Tasks seven and eight describe counting and lists and query for the number of items with a property or a single word answers that would require the creation of a list.

Simple Negation and Indefinite Knowledge Tasks nine and ten have the objective to verify if the AI system can handle natural language constructs in the form of negation and uncertainty.

The tasks are described by a set of statements and questions about them, as can be seen in the boxes below:

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary traveled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts

John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 4: Two Argument Relations

The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden. What is north of the bedroom? A:office
What is the bedroom north of? A:bathroom

Task 5: Three Argument Relations

Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 6: Yes/No Questions

John moved to the playground.
Daniel went to the bathroom.
John went to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Task 7: Counting

Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 8: Lists/Sets

Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

Task 9: Simple Negation

Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Task 10: Indefinite Knowledge

John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

4.3 Results

This section will present our results, including a definition of a scoring system and the ranking for all the tests we executed.

4.3.1 Scoring System

Our implementation of the ToM model outputs textual representations of the mental state of an agent (*Beliefs*), as it could be perceived by an *Observer*. This leads to the question: How can we evaluate the accuracy of our system?

The structure of a *Belief* is as follows:

$$< AGENT, BELIEVES|KNOWS, OBJECT, AFFORDANCE, TARGET\ OBJECT >$$

Beliefs, as can be seen from the description, can be composed only if the AI can achieve the following steps under test:

- Identification of the Agents in the Scene, let us call it **AGT**: Properly identifying how many and what are the current agents in the scene.
- Identification of the Objects in the Scene, **OBJ**: Properly identifying how many and what are the current objects in the scene.
- Interaction, **INT**: Whether the system is capable of understanding the interaction of the objects in the scene and attribute modifications to the environment due to these interactions.
- Sufficiency, **SUFF**: Whether the beliefs output by the system provides information enough for an intelligent system to be able to reply to the questions posed by the environment under test.

As an example, let us analyze the first step of the canonical false-belief task as can be seen in Figure 4.1, and what would be the desirable outcomes of the mental state from a computational system based on our scoring system.

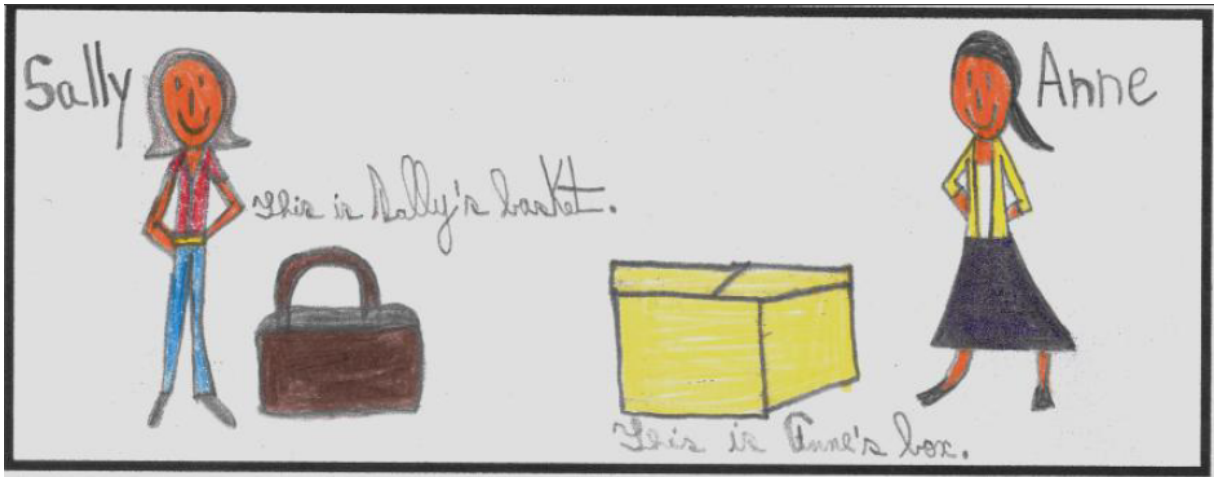

Figure 4.1 – The first step for the Sally-Anne test for false-belief, a mechanism to identify theory of mind capabilities in children.

One example of *Belief* that the computational system should be able to output by analyzing the scene would be of the form:

$$< SALLY, BELIEVES, BASKET, CONTAINS, NONE >$$

Analyzing the outcomes of this *Belief*, we conclude the following:

- **AGT**: The computational system would need to identify that there are a total of two agents in the scene, Sally and Anne, besides the Observer entity. Our *Belief* has correctly identified Sally in the scene.

- **OBJ**: The computational system would need to identify that there are two objects in the scene, Sally's basket, and Anne's box. Our *Belief* has adequately identified the Basket as one of the objects in the scene.
- **INT**: The computational system should be able to attach properties to the objects in the scene. Our *Belief* has properly attached the "CONTAINS" property to the Basket.
- **SUFF**: The computational system should be able to provide enough information for an intelligent system to be able to reply to the questions posed by the environment under test. In this case, our *Belief* has created a statement that Sally knows that the Basket may contain something, which is correct.

It becomes clear that we need the four elements of the construction of one belief, **AGT, OBJ, INT, and SUFF** to be computed correctly in order to create a proper representation. How to rank these in order of importance?

Agents and Objects are necessary to create a scene description. However, only identifying them is not going to take us too far - it will not allow us to create a theory about what happened in the scene and assign *Beliefs* as it is our end objective. Therefore, we should assign lower weights to both these elements. Since we are planning to assign a rank out of 10, it is sensible to limit the ranking for the identification to both at around 40% completion, thus setting for elements the weight of 2 out of 10 each.

Interaction and Sufficiency, on the other hand, are where we should focus our attention. **Interaction** provides us with a measure of the properties of an object in a scene and how agents could interact with it. Interaction is wholly related to the knowledge of an AI about the world and the concept of affordances. It is only natural, in our view, to assign a heavier weight to this element, setting it to 3 out of 10.

Sufficiency is an overall measure of the quality of the AI system under test. Are the *Beliefs* the system is outputting providing the required information we need to qualify what is happening in a scene correctly? This element integrates all the information provided by the primary elements and should be assigned a heavier weight, setting it to 3 out of 10. Summing up, all the four elements weights described here add to ten. This measure will allow us to qualify how well our system is performing.

Alternatives A pause for reflection here is in order before we proceed to evaluation. What are alternative methodologies of measurement of efficiency of an AI system for the task at hand? If we want to measure if someone or some AI has a theory of mind, it either passes or fails; there is no middle ground here. Therefore, we can only rate how close to reproducing a human evaluation we can get to.

4.3.2 Test Case Scores

In this section, we will outline the process for implementing the inputs for each of the test cases under study and the outcomes and scoring results.

Canonical False-Belief Task is the base test case for our system. We started by simulating a visual system (GRASSIOTTO; COSTA, 2020) that would be capable of generating tables 4.2a, 4.2b, 4.2c, 4.2d and 4.2e (the tables show just the first two mindsteps, as can be seen in figure 4.2). Also, we simulated a system that is capable of, given an object, returning specific affordances of the object, as in Table 4.2d.

Figure 4.2 – Sequence of the first two steps for the Sally-Anne test for false-belief, a mechanism to identify theory of mind capabilities in children.

Table 4.2 – Input tables for the canonical false-belief test

(a) entities.txt

t	Entity	Is_agent
1	Sally	True
1	Anne	True
1	Basket	False
1	Box	False
1	Ball	False
2	Sally	True
2	Anne	True
2	Basket	False
2	Box	False
2	Ball	False

(b) positioning.txt

t	Entity	Location
1	Sally	Room
1	Anne	Room
1	Basket	Room
1	Box	Room
1	Ball	Room
2	Sally	Room
2	Anne	Room
2	Basket	Room
2	Box	Room
2	Ball	Room

(c) eye_directions.txt

t	Agent	Object
1	Sally	Anne
1	Sally	Basket
1	Sally	Box
1	Sally	Ball
1	Anne	Sally
1	Anne	Basket
1	Anne	Box
1	Anne	Ball
2	Sally	Anne
2	Sally	Basket
2	Sally	Box
2	Sally	Ball
2	Anne	Sally
2	Anne	Ball
2	Anne	Basket
2	Anne	Box

(d) affordances.txt

Object	Affordance
Box	Contains
Basket	Contains
Ball	Hides
Anne	Exists
Sally	Exists

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Sally	None	None	None
1	Anne	None	None	None
2	Sally	ReachFor	Ball	None
2	Anne	None	None	None

Test Case 4.3.1: Task 1: Canonical False-Belief Task

```
Simulation running mind step: 1

Sally BELIEVES Anne Exists None
Sally BELIEVES Basket Contains None
Sally BELIEVES Box Contains None
Sally BELIEVES Ball Hides None
Anne BELIEVES Sally Exists None
Anne BELIEVES Basket Contains None
Anne BELIEVES Box Contains None
Anne BELIEVES Ball Hides None
Observer KNOWS Sally IS AT Room
Observer KNOWS Anne IS AT Room
```

```
Observer KNOWS Basket IS AT Room
Observer KNOWS Box IS AT Room
Observer KNOWS Ball IS AT Room
...
Simulation running mind step: 9

Sally BELIEVES Anne Exists None
Sally BELIEVES Basket Contains None
Sally BELIEVES Box Contains None
Sally BELIEVES Ball Hidden In Basket
Anne BELIEVES Sally Exists None
Anne BELIEVES Basket Contains None
Anne BELIEVES Box Contains None
Anne BELIEVES Ball OnHand Of Anne
Observer KNOWS Sally IS AT Room
Observer KNOWS Anne IS AT Outside
Observer KNOWS Basket IS AT Room
Observer KNOWS Box IS AT Room
Observer KNOWS Ball IS AT Room

Simulation ended.
```

Input file *entities.txt* (Table 4.2a) simulate a camera input identifying a scene. The camera system identifies a list of entities in the scene and if the entity is an agent. The column t specifies the mind step for the camera information.

Input file *positioning.txt* (Table 4.2b) simulate a positioning system capable of identifying the location of the agents and objects in the environment. The column t specifies the mind step for the positioning system information.

Input file *eye_directions.txt* (Table 4.2c) simulates a visual system capable of identifying eye direction. The information is provided as a triple $< t, Agent, Object >$ where t is the simulation mind step, agent is the agent name, and object is the entity the agent is looking at.

Input file *affordances.txt* (Table 4.2d) presents "affordances" for each of the entities in the scene. Affordances are an entity's properties that show the possible actions users can take with it. For example, a Box may contain other objects, and a Ball may be hidden. Affordances for this system are immutable properties during the simulation timeline.

Input file *intentions.txt* (Table 4.2e) simulates a camera input identifying a scene that could be achieved with human intention understanding video analysis. The camera system identifies the intention of an agent-based on movement and posture information in the scene. The

column *t* specifies the mind step for the camera information.

Results for the **canonical false-belief test** are provided on 4.3.1. Since Sally was not present in the room while Anne took the ball from the basket and hid it, she still believes the ball is in the basket. Therefore, the system we designed can pass the false-belief task.

Using our scoring system, the data for this test is presented in Table 4.3 below. The ratings for the elements are at maximum because the system was able to identify the agents and objects in the scene correctly; it could identify the interactions correctly, and finally, the set of beliefs produced was enough to reply correctly to the question asked by the test case below, "Where does Sally believe the Ball Is?".

Canonical False-Belief Test

Sally and Anne are in the room. Basket, box, and ball are on the floor.
Sally reaches for the ball.
Sally puts the ball in the basket.
Sally exits the room.
Anne reaches for the basket.
Anne gets the ball from the basket.
Anne puts the ball in the box.
Anne exits the room, and Sally enters.
Sally searches for the ball in the room.
Where does Sally believe the Ball Is? A: Basket

Table 4.3 – Score table for Sally-Anne

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
Sally-Anne	Canonical False-Belief	2.0	2.0	3.0	3.0	10.0

Meta bAbI Task 1 test case consists of a question to identify the location of an agent, given one single supporting task (*Mary traveled to the office*). Input tables are provided on Table 4.4. Again, we have entities, positioning, eye directions, affordances, and intentions for this example. Note here that the affordances Table needs to identify that the agents do possess the *Move* affordance, and the intentions Table adds the *Go* intention to enable tracking movement between locations. This task deals with the location of the agents in the environment. By introducing the concept of Observer beliefs for the location of the agent, the beliefs could be produced correctly on 4.3.2.

Table 4.4 – Input tables for Meta bAbI Task 1

(a) entities.txt

t	Entity	Is_agent
1	Mary	True
1	John	True
2	Mary	True
2	John	True

(b) positioning.txt

t	Entity	Location
1	Mary	Bathroom
1	John	Hallway
2	Mary	Office
2	John	Hallway

(c) eye_directions.txt

t	Agent	Object
1	Mary	John
1	John	Mary
2	Mary	John
2	John	Mary

(d) affordances.txt

Object	Affordance
Mary	Move
John	Move

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Mary	Go	Self	Bathroom
1	John	Go	Self	Hallway
2	Mary	Go	Self	Office
2	John	None	Self	None

Test Case 4.3.2: Task 1: Single Supporting Fact

```
Simulation running mind step: 1

Mary BELIEVES John Exists None
John BELIEVES Mary Exists None
Observer KNOWS Mary IS AT Bathroom
Observer KNOWS John IS AT Hallway

Simulation running mind step: 2

Mary BELIEVES John Exists None
John BELIEVES Mary Exists None
Observer KNOWS Mary IS AT Office
Observer KNOWS John IS AT Hallway

Simulation ended.
```

Scoring data for this test are presented at Table 4.5 below. The ratings for the elements are again at maximum because the system was able to identify the agents and objects in the scene properly; it could identify the interactions properly; and finally, the set of beliefs produced was enough to reply to the question asked by the test case below, due to the use of positioning information.

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary traveled to the office.
Where is Mary? A:office

Table 4.5 – Score table for bAbI1

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI01	Single Supporting Task	2.0	2.0	3.0	3.0	10.0

Meta bAbI Task 2 provides two supporting statements that would have to be chained to answer the question asked of where is the agent and the object. This problem is more straightforward than the previous one, as we do not have movement of agents in the system. This time, our system can reply to the question just by tracking positional data as provided in Table 4.6. Results are provided on 4.3.3.

Table 4.6 – Input tables for Meta bAbI Task 2

(a) entities.txt

t	Entity	Is_agent
1	John	True
1	Bob	True
1	Football	False
2	John	True
2	Football	False

(b) positioning.txt

t	Entity	Location
1	John	Playground
1	Bob	Playground
1	Football	Playground
2	John	Playground
2	Bob	Kitchen
2	Football	Playground

(c) eye_directions.txt

t	Agent	Object
1	John	Bob
1	John	Football
1	Bob	John
1	Bob	Football
2	John	Football

(d) affordances.txt

Object	Affordance
John	Exists
Football	Pickup
Bob	Exists

(e) intentions.txt

t	Agent	Intention	Object	Target
1	John	Pickup	Football	None
2	Bob	Go	Self	Kitchen

Test Case 4.3.3: Task 2: Two Supporting Facts

```
Simulation running mind step: 1

John BELIEVES Bob Exists None
John BELIEVES Football OnHand Of John
Bob BELIEVES John Exists None
```

```
Bob BELIEVES Football Pickup None
Observer KNOWS John IS AT
Playground
Observer KNOWS Bob IS AT
Playground
Observer KNOWS Football IS AT Playground

Simulation running mind step: 2

John BELIEVES Bob Exists None
John BELIEVES Football Pickup None
Bob BELIEVES John Exists None
Bob BELIEVES Football Pickup None
Observer KNOWS John IS AT Playground
Observer KNOWS Bob IS AT Kitchen
Observer KNOWS Football IS AT Playground

Simulation ended.
```

Scoring data for this test are presented at Table 4.7 below. Once more, the system was able to identify the agents and objects in the scene properly; it could identify the interactions properly; and finally, the set of beliefs produced was enough to reply to the question asked by the test case below, due to the use of positioning information.

Task 2: Two Supporting Facts

John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Table 4.7 – Score table for bAbI2

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI02	Two Supporting Tasks	2.0	2.0	3.0	3.0	10.0

Meta bAbI Task 3 provides statements that are ordered in time, asking a question about the succession of events. For this problem, our system needs additional positioning data provided on Table 4.8 for the transitions in the position of the agent. This task requires a temporal registry of the beliefs created in each simulation step. The system can identify temporal succession by the internal steps of the creation of beliefs on 4.3.4.

Table 4.8 – Input tables for Meta bAbI Task 3

(a) entities.txt

t	Entity	Is_agent
1	John	True
1	Apple	False
2	John	True
2	Apple	False
3	John	True
3	Apple	False
4	John	True
4	Apple	True

(b) positioning.txt

t	Entity	Location
1	John	Room
1	Apple	Office
2	John	Office
2	Apple	Office
3	John	Kitchen
3	Apple	Kitchen
4	John	Kitchen
4	Apple	Kitchen

(c) eye_directions.txt

t	Agent	Object
1	John	Apple
2	John	Apple
3	John	Apple
4	John	Apple

(d) affordances.txt

Object	Affordance
John	Exists
Apple	Pickup

(e) intentions.txt

t	Agent	Intention	Object	Target
1	John	Pickup	Apple	None
2	John	Go	Self	Office
3	John	Go	Self	Kitchen
4	John	Drop	Apple	None

Test Case 4.3.4: Task 3: Three Supporting Facts

```
Simulation running mind step: 1

John BELIEVES Apple OnHand Of John
Observer KNOWS John IS AT Room
Observer KNOWS Apple IS AT Room

Simulation running mind step: 2

John BELIEVES Apple Pickup None
Observer KNOWS John IS AT Office
Observer KNOWS Apple IS AT Office

Simulation running mind step: 3

John BELIEVES Apple Pickup None
Observer KNOWS John IS AT Kitchen
Observer KNOWS Apple IS AT Kitchen

Simulation running mind step: 4
```

```
John BELIEVES Apple Dropped None
Observer KNOWS John IS AT Kitchen
Observer KNOWS Apple IS AT Kitchen

Simulation ended.
```

Scoring data for this test are presented at Table 4.9 below. This time, the Sufficiency element was rated at zero since there was not enough information in the beliefs to qualify the scene, due to the lack of temporal information to reply to the test case question below - “Where was the apple before the kitchen?”.

Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Table 4.9 – Score table for bAbI3

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI03	Three Supporting Tasks	2.0	2.0	3.0	0.0	7.0

Meta bAbI Task 4 provides statements that qualify the layout of the environment where the agents would be situated. There are no agents or objects described in the test case, so our input tables at Table 4.10 are only describing an *Observer* in the office space. Beliefs are returned on 4.3.5.

Table 4.10 – Input tables for Meta bAbI Task 4

(a) entities.txt

t	Entity	Is_agent
1	Observer	True

(b) positioning.txt

t	Entity	Location
1	Observer	Office

(c) eye_directions.txt

t	Agent	Object
1	Observer	None

(d) affordances.txt

Object	Affordance
Observer	Exists

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Observer	None	None	None

Test Case 4.3.5: Task 4: Two Argument Relations

```
Simulation running mind step: 1

Observer BELIEVES None None None
Observer KNOWS Observer IS AT Office

Simulation ended.
```

Scoring data for this test are presented at Table 4.11 below. Again, the Sufficiency element was rated at zero since there was not enough information in the beliefs to reply to the test case question below, which asks the layout of the rooms in the environment.

Task 4: Two Argument Relations

The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden. What is north of the bedroom? A:office
What is the bedroom north of? A:bathroom

Table 4.11 – Score table for bAbI4

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI04	Two Argument Relations	2.0	2.0	3.0	0.0	7.0

Meta bAbI Task 5 provides three-argument relations and questions the ownership of an object. For this problem, our system needs to understand the property of an object to be transferrable to another agent, as can be seen in Table 4.12 where affordances are defined. This task describes a situation in which one of the agents gives objects (*Cake, Milk*) to another. The system could cope with the input due to the introduction of affordances identifying the transferability of the objects on 4.3.6.

Table 4.12 – Input tables for Meta bAbI Task 5 (Shortened)

(a) entities.txt

t	Entity	Is_agent
3	Mary	True
3	Fred	True
3	Jeff	True
3	Bill	True
3	Cake	False
3	Milk	False

(b) positioning.txt

t	Entity	Location
3	Mary	Room
3	Fred	Room
3	Jeff	Room
3	Bill	Room
3	Cake	Room
3	Milk	Room

(c) eye_directions.txt

t	Agent	Object
3	Mary	Fred
3	...	...
3	Bill	Fred
3	Bill	Jeff
3	Bill	Mary
3	Bill	Cake
3	Bill	Milk

(d) affordances.txt

Object	Affordance
Mary	Exists
Fred	Exists
Bill	Exists
Jeff	Exists
Cake	Transferable
Milk	Transferable

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Mary	Give	Cake	Fred
2	Fred	Give	Cake	Bill
3	Bill	Give	Milk	Jeff

Test Case 4.3.6: Task 5: Three Argument Relations

```
Simulation running mind step: 1


Mary BELIEVES Fred Exists None
Mary BELIEVES Jeff Exists None
Mary BELIEVES Bill Exists None
Mary BELIEVES Cake Was Given To Fred
Mary BELIEVES Milk Transferable None
Fred BELIEVES Mary Exists None
Fred BELIEVES Jeff Exists None
Fred BELIEVES Bill Exists None
Fred BELIEVES Cake Transferable None
Fred BELIEVES Milk Transferable None
Jeff BELIEVES Fred Exists None
Jeff BELIEVES Mary Exists None
Jeff BELIEVES Bill Exists None
Jeff BELIEVES Cake Transferable None
Jeff BELIEVES Milk Transferable None
Bill BELIEVES Fred Exists None
Bill BELIEVES Jeff Exists None
```

```
Bill BELIEVES Mary Exists None
Bill BELIEVES Cake Transferable None
Bill BELIEVES Milk Transferable None
Observer KNOWS Mary IS AT Room
Observer KNOWS Fred IS AT Room
Observer KNOWS Jeff IS AT Room
Observer KNOWS Bill IS AT Room
Observer KNOWS Cake IS AT Room
Observer KNOWS Milk IS AT Room

...
Simulation running mind step: 3

Mary BELIEVES Fred Exists None
Mary BELIEVES Jeff Exists None
Mary BELIEVES Bill Exists None
Mary BELIEVES Cake Transferable None
Mary BELIEVES Milk Transferable None
Fred BELIEVES Mary Exists None
Fred BELIEVES Jeff Exists None
Fred BELIEVES Bill Exists None
Fred BELIEVES Cake Transferable None
Fred BELIEVES Milk Transferable None
Jeff BELIEVES Fred Exists None
Jeff BELIEVES Mary Exists None
Jeff BELIEVES Bill Exists None
Jeff BELIEVES Cake Transferable None
Jeff BELIEVES Milk Transferable None
Bill BELIEVES Fred Exists None
Bill BELIEVES Jeff Exists None
Bill BELIEVES Mary Exists None
Bill BELIEVES Cake Transferable None
Bill BELIEVES Milk Was Given To Jeff
Observer KNOWS Mary IS AT Room
Observer KNOWS Fred IS AT Room
Observer KNOWS Jeff IS AT Room
Observer KNOWS Bill IS AT Room
Observer KNOWS Cake IS AT Room
```

```
Observer KNOWS Milk IS AT Room


Simulation ended.
```

Using our scoring system, the data for this test is presented in Table 4.13 below. The ratings for the elements are at maximum because the system could correctly identify the agents and objects in the scene; it could identify the interactions correctly, and finally, the set of beliefs produced was enough to reply to the question asked below.

Task 5: Three Argument Relations

Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Table 4.13 – Score table for bAbI5

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI05	Three Argument Relations	2.0	2.0	3.0	3.0	10.0

Meta bAbI Task 6 is the most straightforward test possible for the ability of a model to answer true or false to a question. For this problem, our system needs positioning data again provided on Table 4.14. Even though the system is not capable of replying to Yes/No questions in its current form, the set of beliefs could be used as input for a specialist QA NLP system as can be seen on 4.3.7.

Table 4.14 – Input tables for Meta bAbI Task 6

(a) entities.txt

t	Entity	Is_agent
1	John	True
1	Daniel	True
2	John	True
2	Daniel	True

(b) positioning.txt

t	Entity	Location
1	John	Playground
1	Daniel	Bathroom
2	John	Hallway
2	Daniel	Bathroom

(c) eye_directions.txt

t	Agent	Object
1	John	Daniel
1	Daniel	John
2	John	Daniel
2	Daniel	John

(d) affordances.txt

Object	Affordance
John	Exists
Daniel	Exists

(e) intentions.txt

t	Agent	Intention	Object	Target
1	John	Go	Self	Playground
1	Daniel	Go	Self	Bathroom
2	John	Go	Self	Hallway
2	Daniel	None	Self	None

Test Case 4.3.7: Task 6: Yes/No Questions

```
Simulation running mind step: 1

John BELIEVES Daniel Exists None
Daniel BELIEVES John Exists None
Observer KNOWS John IS AT Playground
Observer KNOWS Daniel IS AT Bathroom

Simulation running mind step: 2

John BELIEVES Daniel Exists None
Daniel BELIEVES John Exists None
Observer KNOWS John IS AT Hallway
Observer KNOWS Daniel IS AT Bathroom

Simulation ended.
```

Scoring data for this test are presented at Table 4.15 below. For this test case, we considered that the Sufficiency element could not be rated at maximum since there was not enough information in the beliefs to reply to the test case question that asks a Yes/No question below.

Task 6: Yes/No Questions

John moved to the playground.
Daniel went to the bathroom.
John went to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Table 4.15 – Score table for bAbI6

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI06	Yes/No Questions	2.0	2.0	3.0	1.5	8.5

Meta bAbI Task 7 is a reasonably specific test to check the counting ability of an AI system. For this problem, our input data is provided in Table 4.16. Even though the system is not capable of counting objects in their current form, the set of beliefs could be used as input for a specialist QA NLP system as can be seen on 4.3.8.

Table 4.16 – Input tables for Meta bAbI Task 7

(a) entities.txt

t	Entity	Is_agent
1	Daniel	True
1	Football	False
1	Milk	False
1	Apple	False
2	Daniel	True
2	Football	False
2	Milk	False
2	Apple	False
3	Daniel	True
3	Football	False
3	Milk	False
3	Apple	False
4	Daniel	True
4	Football	False
4	Milk	False
4	Apple	False

(b) positioning.txt

t	Entity	Location
1	Daniel	Room
1	Football	Room
1	Milk	Room
1	Apple	Room
2	Daniel	Room
2	Football	Room
2	Milk	Room
2	Apple	Room
3	Daniel	Room
3	Football	Room
3	Milk	Room
3	Apple	Room
4	Daniel	Room
4	Football	Room
4	Milk	Room
4	Apple	Room

(c) eye_directions.txt

t	Agent	Object
1	Daniel	Football
1	Daniel	Milk
1	Daniel	Apple
2	Daniel	Football
2	Daniel	Milk
2	Daniel	Apple
3	Daniel	Football
3	Daniel	Milk
3	Daniel	Apple
4	Daniel	Football
4	Daniel	Milk
4	Daniel	Apple

(d) affordances.txt

Object	Affordance
Daniel	Exists
Football	Pickup
Football	Drop
Milk	Pickup
Apple	Pickup

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Daniel	Pickup	Football	None
2	Daniel	Drop	Football	None
2	Daniel	Get	Milk	None
2	Daniel	Take	Apple	None

Test Case 4.3.8: Task 7: Counting

```
Simulation running mind step: 1

Daniel BELIEVES Football OnHand Of Daniel
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
```

```
Observer KNOWS Apple IS AT Room

Simulation running mind step: 2

Daniel BELIEVES Football Dropped None
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
Observer KNOWS Apple IS AT Room

Simulation running mind step: 3

Daniel BELIEVES Football Pickup None
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
Observer KNOWS Apple IS AT Room

Simulation running mind step: 4

Daniel BELIEVES Football Pickup None
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
Observer KNOWS Apple IS AT Room

Simulation ended.
```

Scoring data for this test are presented at Table 4.17 below. We considered that both the Interactivity and Sufficiency elements could not be rated at maximum for this test case. The reasoning for that is that the system does not support more than one object associated with an agent in this initial implementation or counting operations.

Task 7: Counting

Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Table 4.17 – Score table for bAbI7

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI07	Counting	2.0	2.0	0.0	0.0	4.0

Meta bAbI Task 8 is quite similar to the previous test, this time to check the ability to compose sets. For this problem, our input data is provided in Table 4.18 Even though the system is not capable of grouping objects in its current form, the set of beliefs could be again used as input for a specialist QA NLP system as can be seen on 4.3.9.

Table 4.18 – Input tables for Meta bAbI Task 8

(a) entities.txt

t	Entity	Is_agent
1	Daniel	True
1	Football	False
1	Milk	False
1	Apple	False
2	Daniel	True
2	Football	False
2	Milk	False
2	Apple	False
3	Daniel	True
3	Football	False
3	Milk	False
3	Apple	False
4	Daniel	True
4	Football	False
4	Milk	False
4	Apple	False

(b) positioning.txt

t	Entity	Location
1	Daniel	Room
1	Football	Room
1	Milk	Room
1	Apple	Room
2	Daniel	Room
2	Football	Room
2	Milk	Room
2	Apple	Room
3	Daniel	Room
3	Football	Room
3	Milk	Room
3	Apple	Room
4	Daniel	Room
4	Football	Room
4	Milk	Room
4	Apple	Room

(c) eye_directions.txt

t	Agent	Object
1	Daniel	Football
1	Daniel	Milk
1	Daniel	Apple
2	Daniel	Football
2	Daniel	Milk
2	Daniel	Apple
3	Daniel	Football
3	Daniel	Milk
3	Daniel	Apple
4	Daniel	Football
4	Daniel	Milk
4	Daniel	Apple

(d) affordances.txt

Object	Affordance
Daniel	Exists
Football	Pickup
Newspaper	Drop
Milk	Pickup
Apple	Pickup

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Daniel	Pickup	Football	None
2	Daniel	Drop	Football	None
2	Daniel	Get	Milk	None
2	Daniel	Take	Apple	None

Test Case 4.3.9: Task 8: Lists/Sets

```
Simulation running mind step: 1

Daniel BELIEVES Football OnHand Of Daniel
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
```

```
Observer KNOWS Apple IS AT Room

Simulation running mind step: 2

Daniel BELIEVES Football Dropped None
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
Observer KNOWS Apple IS AT Room

Simulation running mind step: 3

Daniel BELIEVES Football Pickup None
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
Observer KNOWS Apple IS AT Room

Simulation running mind step: 4

Daniel BELIEVES Football Pickup None
Daniel BELIEVES Milk Pickup None
Daniel BELIEVES Apple Pickup None
Observer KNOWS Daniel IS AT Room
Observer KNOWS Football IS AT Room
Observer KNOWS Milk IS AT Room
Observer KNOWS Apple IS AT Room

Simulation ended.
```

Scoring data for this test are presented in Table 4.19 below. Again, we have considered that both the Interactivity and Sufficiency elements could not be rated at maximum. The reasoning for that is that the system does not support more than one object associated with an agent in this initial implementation or grouping operations.

Task 8: Lists/Sets

Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

Table 4.19 – Score table for bAbI8

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI08	List/Sets	2.0	2.0	0.0	0.0	4.0

Meta bAbI Task 9 is quite similar to Task 1 in the sense that the objective is to track the location of agents in the environment. The unique feature of this test is to describe it through negation statements, which does not affect the setup of our visual system. For this problem, our input data is provided on Table 4.20 and the set of beliefs can be seen on 4.3.10.

Table 4.20 – Input tables for Meta bAbI Task 9

(a) entities.txt

t	Entity	Is_agent
1	Sandra	True
1	Fred	True

(b) positioning.txt

t	Entity	Location
1	Sandra	Office
1	Fred	Elsewhere

(c) eye_directions.txt

t	Agent	Object
1	Sandra	None
1	Fred	None

(d) affordances.txt

Object	Affordance
Sandra	Exists
Fred	Exists

(e) intentions.txt

t	Agent	Intention	Object	Target
1	Sandra	Go	Self	Office
1	Fred	Go	Self	Elsewhere

Test Case 4.3.10: Task 9: Simple Negation

```
Simulation running mind step: 1

Sandra BELIEVES None None None
Fred BELIEVES None None None
Observer KNOWS Sandra IS AT Office
Observer KNOWS Fred IS AT Elsewhere

Simulation ended.
```

Scoring data for this test are presented in Table 4.21 below. Once more, the system could identify the agents and objects in the scene correctly; it could identify the interactions properly; finally, the set of beliefs produced was enough to reply to the question asked by the test case below due to positioning information.

Task 9: Simple Negation

Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Table 4.21 – Score table for bAbI9

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI09	Simple Negation	2.0	2.0	3.0	3.0	10.0

Meta bAbI Task 10 repeats the tracking of agents on the environment, but this time provides insufficient information to the AI system ("Classroom **OR** Playground"). The input data is provided on Table 4.22 and the set of beliefs can be seen on 4.3.11. The system as designed cannot define uncertainty, so the input was modified to include as the location in the positioning data two possibilities.

Table 4.22 – Input tables for Meta bAbI Task 10

(a) entities.txt

t	Entity	Is_agent
1	John	True
1	Sandra	True

(b) positioning.txt

t	Entity	Location
1	John	"Classroom Or Playground"
1	Sandra	Garden

(c) eye_directions.txt

t	Agent	Object
1	John	None
1	Sandra	None

(d) affordances.txt

Object	Affordance
John	Exists
Sandra	Exists

(e) intentions.txt

t	Agent	Intention	Object	Target
1	John	None	None	None
1	Sandra	None	None	None

Test Case 4.3.11: Task 10: Indefinite Knowledge

```
Simulation running mind step: 1

John BELIEVES None None None
Sandra BELIEVES None None None
Observer KNOWS John IS AT Classroom Or Playground
```

```
Observer KNOWS Sandra IS AT Garden


Simulation ended.
```

Scoring data for this test are presented at Table 4.23 below. Once more, the system could identify the agents and objects in the scene properly; it could identify the interactions properly, but the information output cannot by itself handle uncertainty, so the Sufficiency element had to be set at a lower score.

Task 10: Indefinite Knowledge

John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

Table 4.23 – Score table for bAbI10

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
FB bAbI10	Indefinite Knowledge	2.0	2.0	3.0	1.5	8.5

4.3.3 Overall Score Results

The Table 4.24 provides the scores for the execution of the 11 test cases we described in section 4.2.

Table 4.24 – Score table for all test cases

Test Case	Description	AGT	OBJ	INT	SUFF	TOTAL
Sally-Anne	Canonical False-Belief	2.0	2.0	3.0	3.0	10.0
FB bAbI01	Single Supporting Task	2.0	2.0	3.0	3.0	10.0
FB bAbI02	Two Supporting Tasks	2.0	2.0	3.0	3.0	10.0
FB bAbI03	Three Supporting Tasks	2.0	2.0	3.0	0.0	7.0
FB bAbI04	Two Argument Relations	2.0	2.0	3.0	0.0	7.0
FB bAbI05	Three Argument Relations	2.0	2.0	3.0	0.0	10.0
FB bAbI06	Yes/No Questions	2.0	2.0	3.0	1.5	8.5
FB bAbI07	Counting	2.0	2.0	0.0	0.0	4.0
FB bAbI08	List/Sets	2.0	2.0	0.0	0.0	4.0
FB bAbI09	Simple Negation	2.0	2.0	3.0	3.0	10.0
FB bAbI10	Indefinite Knowledge	2.0	2.0	3.0	1.5	8.5
AVERAGE						**9.00**

We note that we managed, on average, to maintain a high score for the test cases under evaluation. The system we designed is reasonably adaptable to the range of test cases we tested, even though it has been targeted at solving the Canonical False-Belief Test. There are weaknesses to be addressed, including analyzing environments where agents and objects are not present, introducing more sophisticated NLP systems to handle question answering, and adding further flexibility to the architecture to handle the multiplicity of objects and affordances.

4.4 Discussion

In this chapter, we presented the simulation results for some scenarios.

We described our methodology and discussed alternative approaches for systems designed to understand the world, including video understanding, visual question answering, and physical and mechanical reasoning. We debated, briefly, how these approaches differ from ours. We considered then simulating the canonical false-belief task and verified that the system we designed could reproduce the outcomes of a human mind.

After that, we proceeded to analyze a dataset proposed by Meta research to validate the capabilities of an artificial intelligence system. We selected several test tasks from the dataset, created input data, and fed this data to the system we designed. We noted that our proposed cognitive architecture could handle the test tasks and reproduce the possible outcomes according to expectations.

The next chapter will discuss the work we have proposed and consider the following steps to advance the state-of-the-art in this field.

5 Conclusions

5.1 Final Remarks

Our research question wanted to know what it would take to implement theory of mind in a computational system. Throughout this work, we found out that we could create software components, or modules, that could simulate models from psychology. Employing flexible software components, we managed to implement it using the CST Toolkit, and we had success in achieving our objectives.

In this process, we created the CogToM cognitive architecture. It was designed as a platform to validate this viability of a computational system to implement a model of the theory of mind. The initial drive for this work had been the implementation of an AI capable of assisting people in the autism spectrum with the understanding of the environment, including agents, object interaction, and intentions. What is quite interesting is that we have undoubtedly surpassed our initial goals.

In order to achieve the AI we wanted, we looked for psychological models of the mind that we could implement to imitate the human cognitive apparatus. We selected the mindreading model from the literature as proposed by Simon Baron-Cohen. There are some alternative theories for explaining autism deficits, such as the Empathising-Systemizing theory, deficits in executive functioning of the brain, and others. The mindreading model seemed to be an excellent match to what we wanted to achieve due to the availability of a well-established model in the literature (BARON-COHEN, 2009; FRITH, 1996).

We established a target for our AI system: the capability of adequately solving the problem posed by false-belief tasks understanding. If a system could do that, we would be on the right path to create the type of robotic intelligence we wanted.

The psychological models we found in the literature do not provide a complete basis for implementing the cognitive architecture. In the case of the mindreading model, we found that it lacks a more profound analysis of how the mind creates and maintains data about the world.

We noticed that during the analysis of the computational system. We had to augment the model, adding the ability to assign properties to objects and understand human intention. We need more information about the world, both in terms of common world knowledge, with constructs from semantic memory in the form of affordances, and episodic information about the world in positioning and human intentions.

When implementing the architecture using the CST toolkit, we noted that the flexibility afforded and the concepts of codelets and memory objects, led to a fine mapping of the

psychological model we identified. However, our model is sequential, whereas the CST Toolkit is designed for parallel processing. Since the outcomes of the initial modules have to be fed to the next module in the architecture, we had to make do with implementation of blocking and unblocking constructs using memory objects. Alternatively, this same behavior could have been implemented using the conscience mechanism from the CST toolkit.

The architecture we implemented was capable of passing the canonical false-belief task as defined by researchers in the autism spectrum. The system was designed in such a way to be generic enough to allow for testing with a set of tasks usually used to qualify the capabilities of an artificial intelligence system, the Meta bAbI dataset.

We created and described a scoring system to qualify the outputs provided by the system. This scoring system showed us good results with the set of tasks from the bAbl dataset.

Although designed as an assistant for people in the autism spectrum, CogToM uses text processing at its core for the production of beliefs. We see that it may find applications in the domain of natural language processing research and cognitive modeling in general.

It is fair to say that the CogToM cognitive architecture holds promise as a base system on which future assistive systems for people in the autism spectrum can be based.

5.2 Future Work

As an immediate follow-up to this work, we believe that the model of the theory of mind we implemented could be considered a basis for future research. In order to achieve that, the set of components we created using the CST toolkit could be made available to the community. The modeling for the mindreading model could be augmented to include other scenarios, including constructs for beliefs we did not consider in this initial implementation, such as models for the mind states of pretending, thinking, knowing, believing, and so on.

In this initial plan, we also also consider that reinforcement learning techniques could be used to reinforce the system's outputs on human input. This would allow for the system to qualify the choice of the *beliefs* according to these techniques.

As future directions for this work, we consider symbolic (and the new Neuro-Symbolic) AI to be considered an approach for modeling the mind constructs, so this should be approached as a new avenue for the research into models for the theory of mind.

Finally, it is worth noting that the theory of mind we studied here is but one of many mind constructs. Other theories seek to explain the deficits associated with functionalities of the brain, such as executive function or mirror systems. We believe further research in these systems, and implementation in computational systems, are a worthwhile pursuit.

Bibliography

ANDERSON, J. R.; LEBIERE, C. J. *The atomic components of thought*. [S.l.]: Psychology Press, 2014. Cited at page 25.

ANDERSON, J. R.; MATESSA, M.; LEBIERE, C. Act-r: A theory of higher level cognition and its relation to visual attention. *Human–Computer Interaction*, Taylor & Francis, v. 12, n. 4, p. 439–462, 1997. Cited at page 25.

ANTOL, S. et al. Vqa: Visual question answering. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2015. p. 2425–2433. Cited at page 45.

ASPERGER, H.; FRITH, U. T. 'autistic psychopathy'in childhood. Cambridge University Press, 1991. Cited at page 19.

ATKINSON, R. C.; SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In: *Psychology of learning and motivation*. [S.l.]: Elsevier, 1968. v. 2, p. 89–195. Cited at page 31.

AUGUSTYN, M. Autism spectrum disorder: terminology, epidemiology, and pathogenesis. *UpToDate. Waltham MA.*, 2019. Cited at page 14.

BAARS, B. J.; FRANKLIN, S. An architectural model of conscious and unconscious brain functions: Global workspace theory and ida. *Neural networks*, Elsevier, v. 20, n. 9, p. 955–961, 2007. Cited at page 26.

BAARS, B. J.; FRANKLIN, S. Consciousness is computational: The lida model of global workspace theory. *International Journal of Machine Consciousness*, World Scientific, v. 1, n. 01, p. 23–32, 2009. Cited at page 26.

BADDELEY, A.; EYSENCK, M. W.; ANDERSON, M. C. *Memory*. 3rd. ed. [S.l.]: Psychology Press, 2014. Cited at page 31.

BARON-COHEN, S. Precursors to a theory of mind: Understanding attention in others. *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, v. 1, p. 233–251, 1991. Cited at page 30.

BARON-COHEN, S. *Mindblindness: An essay on autism and theory of mind*. [S.l.]: MIT press, 1997. Cited 6 times at pages 8, 13, 21, 23, 30, and 35.

BARON-COHEN, S. The empathising-systemising theory of autism: implications for education. *Tizard Learning Disability Review*, Emerald Group Publishing Limited, 2009. Cited at page 76.

BARON-COHEN, S. et al. Does the autistic child have a "theory of mind". *Cognition*, v. 21, n. 1, p. 37–46, 1985. Cited 6 times at pages 8, 17, 21, 22, 45, and 47.

BARRETT, H. C.; KURZBAN, R. Modularity in cognition: framing the debate. *Psychological review*, American Psychological Association, v. 113, n. 3, p. 628, 2006. Cited at page 20.

BATTAGLIA, P. W.; HAMRICK, J. B.; TENENBAUM, J. B. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 110, n. 45, p. 18327–18332, 2013. Cited at page 45.

BLAKEMORE, S.-J.; DECETY, J. From the perception of action to the understanding of intention. *Nature reviews neuroscience*, Nature Publishing Group, v. 2, n. 8, p. 561–567, 2001. Cited at page 28.

BONCHEK-DOKOW, E.; KAMINKA, G. A. Towards computational models of intention detection and intention prediction. *Cognitive Systems Research*, Elsevier, v. 28, p. 44–79, 2014. Cited at page 28.

BOUCENNA, S. et al. Interactive technologies for autistic children: A review. *Cognitive Computation*, Springer, v. 6, n. 4, p. 722–740, 2014. Cited at page 15.

BUESCHER, A. V. et al. Costs of autism spectrum disorders in the united kingdom and the united states. *JAMA pediatrics*, American Medical Association, v. 168, n. 8, p. 721–728, 2014. Cited at page 14.

CARRUTHERS, P. On fodor's problem. *Mind & Language*, Wiley Online Library, v. 18, n. 5, p. 502–523, 2003. Cited at page 20.

CASSIMATIS, N. L. *Polyscheme: a cognitive architecture for integrating multiple representation and inference schemes*. Tese (Doutorado) — Massachusetts Institute of Technology, 2001. Cited at page 24.

CHARNIAK, E.; GOLDMAN, R. P. A bayesian model of plan recognition. *Artificial Intelligence*, Elsevier, v. 64, n. 1, p. 53–79, 1993. Cited at page 28.

CHOMSKY, N. Language and other cognitive systems. what is special about language? *Language learning and development*, Taylor & Francis, v. 7, n. 4, p. 263–278, 2011. Cited at page 20.

CHOMSKY, N. The galilean challenge. *Inference: International Review of Science*, v. 3, n. 1, 2017. Cited at page 20.

DAUTENHAHN, K.; WERRY, I. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, John Benjamins, v. 12, n. 1, p. 1–35, 2004. Cited at page 15.

DICKSTEIN-FISCHER, L. A. et al. Socially assistive robots: current status and future prospects for autism interventions. *Innovation and Entrepreneurship in Health*, Dove Press, v. 5, p. 15–25, 2018. Cited at page 15.

FIRTH, R. Ethical absolutism and the ideal observer. *Philosophy and Phenomenological Research*, JSTOR, v. 12, n. 3, p. 317–345, 1952. Cited at page 28.

FLAVELL, J. H. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, American Psychological Association, v. 34, n. 10, p. 906, 1979. Cited at page 24.

FLAVELL, J. H. Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly (1982-)*, JSTOR, p. 274–290, 2004. Cited 2 times at pages 20 and 21.

FODOR, J. A. *The modularity of mind*. [S.l.]: MIT press, 1983. Cited 2 times at pages 19 and 20.

FRITH, U. Cognitive explanations of autism. *Acta Paediatrica*, Wiley Online Library, v. 85, p. 63–68, 1996. Cited at page 76.

FRITH, U. *Autism: Explaining the enigma*. [S.l.]: Blackwell Publishing, 2003. Cited at page 14.

GALLISTEL, C. R. Prelinguistic thought. *Language learning and development*, Taylor & Francis, v. 7, n. 4, p. 253–262, 2011. Cited at page 20.

GAN, Z. et al. Semantic compositional networks for visual captioning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 5630–5639. Cited at page 45.

GIBSON, J. J. *The ecological approach to visual perception: classic edition*. [S.l.]: Psychology Press, 2014. Cited at page 27.

GIUBILINI, A.; SAVULESCU, J. The artificial moral advisor. the "ideal observer" meets artificial intelligence. *Philosophy & technology*, Springer, v. 31, n. 2, p. 169–188, 2018. Cited 2 times at pages 28 and 30.

GOERTZEL, B. et al. A world survey of artificial brain projects, part ii: Biologically inspired cognitive architectures. *Neurocomputing*, Elsevier, v. 74, n. 1-3, p. 30–49, 2010. Cited at page 24.

GOLDMAN, A. *2012. Theory of Mind*. [S.l.]: The Oxford handbook of philosophy of cognitive science, 2012. Cited at page 20.

GRANDIN, T. *Thinking in pictures, expanded edition: My life with autism*. [S.l.]: Vintage, 2008. Cited at page 13.

GRANDIN, T. *The way I see it: A personal look at autism & Asperger's*. [S.l.]: Future Horizons, 2011. Cited at page 13.

GRANDIN, T.; PANEK, R. *The autistic brain: Thinking across the spectrum*. [S.l.]: Houghton Mifflin Harcourt, 2013. Cited at page 13.

GRASSIOTTO, F.; COSTA, P. D. P. *CogToM-CST Source Code*. 2020. <https://github.com/AI-Unicamp/CogTom-cst/>. Last checked on Aug 27, 2021. Available from Internet: <https://github.com/AI-Unicamp/CogTom-cst/>. Cited 3 times at pages 17, 40, and 52.

GRASSIOTTO, F.; COSTA, P. D. P. Cogtom: A cognitive architecture implementation of the theory of mind. In: *ICAART (2)*. [S.l.: s.n.], 2021. p. 546–553. Cited at page 16.

HAUSER, M. D.; CHOMSKY, N.; FITCH, W. T. The faculty of language: what is it, who has it, and how did it evolve? *science*, American Association for the Advancement of Science, v. 298, n. 5598, p. 1569–1579, 2002. Cited at page 20.

HENDRICKS, L. A. et al. Localizing moments in video with natural language. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 5803–5812. Cited at page 45.

HOLTZEN, S. et al. Inferring human intent from video by sampling hierarchical plans. In: IEEE. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. [S.l.], 2016. p. 1489–1496. Cited at page 27.

JALIAAWALA, M. S.; KHAN, R. A. Can autism be catered with artificial intelligence-assisted intervention technology? a comprehensive survey. *Artificial Intelligence Review*, Springer, v. 53, n. 2, p. 1039–1069, 2020. Cited at page 15.

KAMINKA, G. A. Curing robot autism: a challenge. In: *AAMAS*. [S.l.: s.n.], 2013. p. 801–804. Cited at page 14.

KANNER, L. et al. Autistic disturbances of affective contact. *Nervous child*, New York, v. 2, n. 3, p. 217–250, 1943. Cited at page 19.

KAWALL, J. Ideal observer theories. *International Encyclopedia of Ethics*, Blackwell Publishing Ltd Oxford, UK, 2013. Cited at page 28.

KLIN, A. Autism and asperger syndrome: an overview. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 28, p. s3–s11, 2006. Cited 2 times at pages 14 and 19.

KOTSERUBA, I.; TSOTSOS, J. K. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, Springer, p. 1–78, 2018. Cited at page 24.

LAVELLE, J. S. Theory-theory and the direct perception of mental states. *Review of Philosophy and Psychology*, Springer, v. 3, n. 2, p. 213–230, 2012. Cited 2 times at pages 20 and 21.

LAVELLE, T. A. et al. Economic burden of childhood autism spectrum disorders. *Pediatrics*, American Academy of Pediatrics, v. 133, n. 3, p. e520–e529, 2014. Cited at page 14.

LERER, A.; GROSS, S.; FERGUS, R. Learning physical intuition of block towers by example. In: BALCAN, M.; WEINBERGER, K. Q. (Ed.). *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*. JMLR.org, 2016. (JMLR Workshop and Conference Proceedings, v. 48), p. 430–438. Available from Internet: <http://proceedings.mlr.press/v48/lerer16.html>. Cited at page 45.

LESLIE, A. M. Tomm, toby, and agency: Core architecture and domain specificity. *Mapping the mind: Domain specificity in cognition and culture*, v. 29, p. 119–48, 1994. Cited at page 25.

LIETO, A. et al. *The role of cognitive architectures in general artificial intelligence*. [S.l.]: Elsevier, 2018. Cited at page 24.

LIMA, A. M. O. d. et al. Analysis of softwares for emotion recognition in children and teenagers with autism spectrum disorder. *Revista CEFAC*, SciELO Brasil, v. 21, 2019. Cited at page 15.

LORD, C. et al. Autism spectrum disorders. *Neuron*, Elsevier, v. 28, n. 2, p. 355–363, 2000. Cited at page 19.

MARTINS, M. et al. *Diretrizes de Atenção à Reabilitação da Pessoa com Transtornos do Espectro do Autismo (TEA)*. [S.l.]: Brasília: Ministério da Saúde, 2014. Cited at page 14.

MCCARTHY, J.; HAYES, P. J. Some philosophical problems from the standpoint of artificial intelligence. In: *Readings in artificial intelligence*. [S.l.]: Elsevier, 1981. p. 431–450. Cited at page 43.

MCCLELLAND, T. Ai and affordances for mental action. *environment*, v. 1, p. 127, 2017. Cited at page 27.

META RESEARCH. *Meta Research bAbI project*. 2021. <https://research.fb.com/downloads/babi/>. Last checked on Oct 27, 2021. Available from Internet: <https://research.fb.com/downloads/babi/>. Cited 2 times at pages 43 and 47.

MILLER, G. A. The cognitive revolution: a historical perspective. *Trends in cognitive sciences*, Elsevier, v. 7, n. 3, p. 141–144, 2003. Cited at page 20.

MONTESANO, L. et al. Learning object affordances: from sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, IEEE, v. 24, n. 1, p. 15–26, 2008. Cited at page 27.

PALECEK, M. Modularity of mind: Is it time to abandon this ship? *Philosophy of the Social Sciences*, SAGE Publications Sage CA: Los Angeles, CA, v. 47, n. 2, p. 132–144, 2017. Cited at page 20.

PAPERT, S. Children, computers and powerful ideas. *New York: Basic Books*, v. 10, p. 1095592, 1990. Cited at page 15.

PARAENSE, A. L. et al. The cognitive systems toolkit and the cst reference cognitive architecture. *Biologically Inspired Cognitive Architectures*, Elsevier, v. 17, p. 32–48, 2016. Cited at page 25.

PAULA, C. S. et al. Brief report: prevalence of pervasive developmental disorder in brazil: a pilot study. *Journal of autism and developmental disorders*, Springer, v. 41, n. 12, p. 1738–1742, 2011. Cited at page 14.

PREMACK, D.; WOODRUFF, G. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, Cambridge University Press, v. 1, n. 4, p. 515–526, 1978. Cited 2 times at pages 20 and 21.

PYNADATH, D. V. et al. Modeling two-player games in the sigma graphical cognitive architecture. In: SPRINGER. *International conference on artificial general intelligence*. [S.l.], 2013. p. 98–108. Cited at page 24.

QUILLAN, M. R. *Semantic memory*. [S.l.], 1966. Cited at page 31.

ROSENBLOOM, P. S. The sigma cognitive architecture and system. *AISB Quarterly*, v. 136, p. 4–13, 2013. Cited at page 24.

ROSENBLOOM, P. S.; DEMSKI, A.; USTUN, V. The sigma cognitive architecture and system: Towards functionally elegant grand unification. *Journal of Artificial General Intelligence*, De Gruyter Poland, v. 7, n. 1, p. 1, 2016. Cited at page 24.

ŞAHIN, E. et al. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, Sage Publications Sage UK: London, England, v. 15, n. 4, p. 447–472, 2007. Cited at page 27.

SALLY, D.; HILL, E. The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of economic psychology*, Elsevier, v. 27, n. 1, p. 73–97, 2006. Cited at page 21.

SAVULESCU, J.; MASLEN, H. Moral enhancement and artificial intelligence: moral ai? In: *Beyond Artificial Intelligence*. [S.l.]: Springer, 2015. p. 79–95. Cited 2 times at pages 28 and 30.

SCASSELLATI, B. Theory of mind for a humanoid robot. *Autonomous Robots*, Springer, v. 12, n. 1, p. 13–24, 2002. Cited at page 25.

SCASSELLATI, B. M. *Foundations for a Theory of Mind for a Humanoid Robot*. Tese (Doutorado) — Massachusetts Institute of Technology, 2001. Cited at page 25.

SCHAAFSMA, S. M. et al. Deconstructing and reconstructing theory of mind. *Trends in cognitive sciences*, Elsevier, v. 19, n. 2, p. 65–72, 2015. Cited at page 21.

SHAPIRO, L. A. The routledge handbook of embodied cognition. Routledge New York, 2014. Cited at page 29.

SPAULDING, S. Embodied cognition and theory of mind. 2014. Cited at page 28.

STEEDMAN, M. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, Springer, v. 25, n. 5, p. 723–753, 2002. Cited at page 27.

STOFFREGEN, T. A. Affordances as properties of the animal-environment system. *Ecological psychology*, Taylor & Francis, v. 15, n. 2, p. 115–134, 2003. Cited at page 27.

SUBETHA, T.; CHITRAKALA, S. A survey on human activity recognition from videos. In: IEEE. *2016 international conference on information communication and embedded systems (ICICES)*. [S.l.], 2016. p. 1–7. Cited at page 28.

SUN, R. The clarion cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction*, p. 79–99, 2006. Cited at page 26.

TRAFTON, J. G. et al. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, IEEE, v. 35, n. 4, p. 460–470, 2005. Cited at page 24.

TRIONA, L. M.; MASNICK, A. M.; MORRIS, B. J. What does it take to pass the false belief task? an act-r model. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. [S.l.: s.n.], 2002. v. 24, n. 24. Cited at page 25.

TULVING, E. Episodic and semantic memory: Where should we go from here? *Behavioral and Brain Sciences*, Cambridge University Press, v. 9, n. 3, p. 573–577, 1986. Cited at page 32.

TURVEY, M. T. Affordances and prospective control: An outline of the ontology. *Ecological psychology*, Taylor & Francis, v. 4, n. 3, p. 173–187, 1992. Cited at page 27.

VALADÃO, C. T. et al. Analysis of the use of a robot to improve social skills in children with autism spectrum disorder. *Research on Biomedical Engineering*, SciELO Brasil, v. 32, n. 2, p. 161–175, 2016. Cited at page 14.

WESTON, J. et al. Towards ai-complete question answering: A set of prerequisite toy tasks. In: BENGIO, Y.; LECUN, Y. (Ed.). *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. [S.l.: s.n.], 2016. Cited 3 times at pages 17, 46, and 47.

WHO. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. [S.l.]: World Health Organization, 1993. v. 2. Cited at page 14.

YI, K. et al. Clevrer: Collision events for video representation and reasoning. In: *International Conference on Learning Representations*. [S.l.: s.n.], 2019. Cited at page 45.

YU, Z. et al. Human intention understanding based on object affordance and action classification. In: IEEE. *2015 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2015. p. 1–6. Cited at page 27.

ZADEH, A. et al. Social-iq: A question answering benchmark for artificial social intelligence. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 8807–8817. Cited at page 45.