

# CogToM: A Cognitive Architecture Implementation of the Theory of Mind

Fabio Grassiotto<sup>a</sup> and Paula Dorhofer Paro Costa<sup>b</sup>

*Dept. of Comp. Engineering and Industrial Automation (DCA), School of Electrical and Computer Engineering,  
University of Campinas, Campinas, Brazil*

**Keywords:** Autism, Cognitive Architectures, Artificial Intelligence.

**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. In this work, we propose CogToM, a novel cognitive architecture designed to process the output of computer systems and to reason according to the Theory of Mind. In particular, we present a computational implementation for the psychological model of the Theory of Mind proposed by Baron-Cohen and we explore the usefulness of the concepts of Affordances and Intention Detection to augmenting the effectiveness of the proposed architecture. We verify the results by evaluating both a canonical false-belief and a number of the Facebook bAbI dataset tasks.

## 1 INTRODUCTION

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 ranges from 1 in 40 to 1 in 500, according to population and methodology used (Augustyn, 2019).


Surveys on interactive technologies for autism show that the research in the field has had a focus on diagnosis, monitoring, assessment and intervention tools, interactive or virtual environments, mobile and wearable applications, educational devices, games, and other therapeutic devices or systems (Boucenna et al., 2014; Picard, 2009; Kientz et al., 2019; Jali-aawala and Khan, 2020). However, we highlight the lack of computational systems targeted towards individuals with autism with the express intent of assisting them in real-time with their impairments in social interactions.


Our understanding is that these systems should be designed to analyze environmental and visual social cues that are not readily interpreted by those indi-

viduals in the spectrum and provide expert advice on the best alternative for interaction, improving the outcomes of the social integration for these individuals. However, the lack of systems like this can be partly explained by the limitations of the state-of-the-art machine learning models, which are capable of translate languages, recognize objects, and spoken speech, but still struggles to extract the underlying logical, temporal and causal structure from a massive amount of training data. In this scenario, neurosymbolic artificial intelligence have emerged as an approach to the problem. In CLEVRER, for example, the authors propose a Neuro-Symbolic Dynamic Reasoning (NS-DR) model that combines symbolic logic reasoning and neural nets for pattern recognition (Yi\* et al., 2020).

In this work, we also explore the idea of a specialized visual recognition system that could be attached to a reasoning system. In particular, we focus on the modeling of a reasoning system that is inspired by studies on autism, and we explore existing mechanisms that the human mind uses for facilitating social interaction.

There are several approaches for explaining the deficits brought about by ASD, including environmental, genetics and cognitive. In one of the cognitive approaches, Simon Baron-Cohen proposed the mind-blindness theory of autism (Baron-Cohen, 1997). His work proposed the existence of the mindreading sys-

<sup>a</sup>  <https://orcid.org/0000-0003-1885-842X>

<sup>b</sup>  <https://orcid.org/0000-0002-1534-5744>

tem and established that the cognitive delays associated with autism are related to deficits in the development of such a system.

The mindreading system is directly related to the concept of Theory of Mind (ToM) as the innate ability of attributing mental states to oneself and others and to understand beliefs and desires that are distinct from one's own (Premack and Woodruff, 1978).

Later, research has shown that individuals with ASD show deficits in ToM (Kimhi, 2014; Baraka et al., 2019; Baron-Cohen, 2001). The deficits can be demonstrated in a number of test tasks, in particular false-belief tasks, which are test tasks designed to evaluate children's capacity to understand other people's mental states (Baron-Cohen, 1990).

This work proposes and evaluates CogToM, a novel cognitive architecture that models the Theory of Mind, as hypothesized by Baron-Cohen, that process inputs provided by an external specialized visual recognition system. The paper is organized as follows. In Section 2, we review key concepts related to cognitive architectures, false-belief tests and the key-components of the Baron-Cohen mindreading model. Following, in Section 2 we describe our proposal in detail. In Section 6 we evaluate our model, using the bAbI dataset. The paper is concluded in Section 7.

## 2 RELATED WORK

### 2.1 ToM in Cognitive Architectures

Not many cognitive architectures have supported the implementation of the ToM ability up to this point. Sigma has demonstrated an application for simultaneous-move games. Polyscheme explored perspective-taking for robots' interaction with humans. ACT-R has built models of false-belief tasks and later implemented them on a mobile robot.

Brian Scasselatti PhD thesis (Scasselatti, 2001) proposed a novel architecture called "Embodied Theory of Mind" in which he presented psychological theories on the development of ToM in children, discussing the potential application of both in robotics with the purpose of applying psychological models to the detection of human faces and identifying agents.

Sigma, Polyscheme, ACT-R and others proposed integrating principles of the ToM to enable specific robotic behavior to simulate human-like capabilities of social and robotic interaction. However, none of them proposed an *Observer-like implementation* as we seek here with the CogTom cognitive architecture with the express intent of assistance.



Figure 1: The Sally-Anne test for false-belief. Adapted from (Baron-Cohen et al., 1985), drawing by Alice Grassiotto.

### 2.2 Autism and False-Belief Tasks

False-Belief tasks are a type of task used in the study of ToM to check if a child understands that another person does not possess the same knowledge as herself. In (Baron-Cohen et al., 1985), Baron-Cohen and Frith proposed the *Sally-Anne test* (Figure 1) 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.

The results presented on the article supported the hypothesis that autistic children, in general, fail to employ a ToM, due to the inability of representing mental states. It is thought that this lack of predicting ability causes deficits in the social skills for people in the autism spectrum, making it much harder to face the challenges of social interaction.

### 2.3 Mindreading

The mindreading model (Baron-Cohen, 1997) (Figure 2) seeks to understand the human mindreading

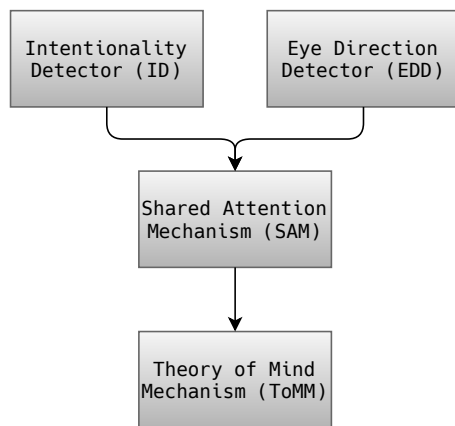


Figure 2: The ToM Model.

process by proposing a set of four separate components. **Intentionality Detector (ID)** is a perceptual device that is able to interpret movement and identify agents from objects and assign goals and desires. The **Eye Direction Detector (EDD)** is a visual system able to 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. The **Shared Attention Mechanism (SAM)** builds internal representations that specify relationships between an agent, the self and a third object. Finally, the **Theory of Mind Mechanism (ToMM)** completes the agent development of mindreading by representing the agent mental states.

### 3 AFFORDANCES

The concept of affordances has been extended as described by (McClelland, 2017) including applications for affordances in robotics (Montesano et al., 2008), (Şahin et al., 2007) as a process to encode the relationships between actions, objects and effects.

For CogToM, the concept of affordances has found use in the analysis of the environment to assign properties to the objects and the environment it is currently situated.

### 4 INTENTION DETECTION

Intention Understanding, a requirement for human-machine interaction (Yu et al., 2015), allows for robots to deduce the possible human intention by considering the relationship between objects and actions.

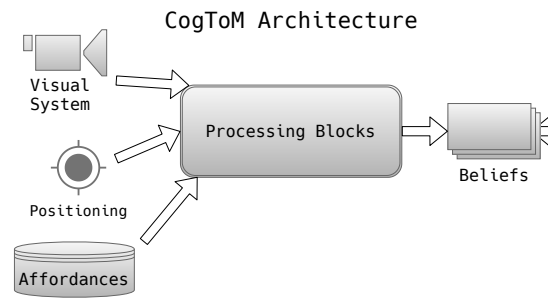


Figure 3: CogToM Cognitive Architecture.

CogToM architecture proposes using external systems capable of understanding human intention to augment the environmental analysis it requires.

## 5 CogToM PROPOSAL

The CogToM architecture (Figure 3) is designed as a central processing unit to implement decision-making functionality with the purpose of implementing an AI *Observer*. This *Observer* has the purpose of passing 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, on the form of a visual camera system capable of identifying agents and objects as well as its locations, eye direction of the agents, and human intention.

Affordances are provided to the system as a database, or rather a dictionary of known properties of the objects in the scene.

The outputs of the system (*Beliefs*) are textual representations of the mental state of an agent as perceived by the *Observer*.

For the purpose of simulating the external visual and affordances system, the required set of inputs are modeled through text files.

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”.

As an example, we will present here the set of inputs the system would require for running two mind steps in the Sally-Anne test on Table 1.

- **Mind Step 1:** Sally and Anne are in the room. Basket, box and ball are on the floor.
- **Mind Step 2:** Sally reaches for the ball.

Input file *entities.txt* (Table 1a) simulate a camera input identifying a scene. The camera system identifies a list of entities in the scene and if the entity is an

Table 1: Input tables for the canonical false-belief test.

(a) entities.txt			(b) positioning.txt			(c) eye_directions.txt			(d) affordances.txt	
t	Entity	Is_agent	t	Entity	Location	t	Agent	Object	Object	Affordance
1	Sally	True	1	Sally	Room	1	Sally	Anne	Box	Contains
1	Anne	True	1	Anne	Room	1	Sally	Basket	Basket	Contains
1	Basket	False	1	Basket	Room	1	Sally	Box	Ball	Hides
1	Box	False	1	Box	Room	1	Sally	Ball	Anne	Exists
1	Ball	False	1	Ball	Room	1	Anne	Sally	Sally	Exists
2	Sally	True	2	Sally	Room	1	Anne	Basket		
2	Anne	True	2	Anne	Room	1	Anne	Box		
2	Basket	False	2	Basket	Room	1	Anne	Ball		
2	Box	False	2	Box	Room	2	Sally	Anne		
2	Ball	False	2	Ball	Room	2	Sally	Basket		
						2	Sally	Box		
						2	Sally	Ball		
						2	Anne	Sally		
						2	Anne	Ball		
						2	Anne	Basket		
						2	Anne	Box		

(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

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

Input file *positioning.txt* (Table 1b) 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 1c) simulates a visual system capable of identifying eye direction. The information is provided as a triple  $\langle t, Agent, Object \rangle$  where *t* is the simulation mind step, agent is agent name and object is the entity the agent is looking at.

Input file *affordances.txt* (Table 1d) 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 the purpose of this system, are immutable properties during the simulation timeline.

Input file *intentions.txt* (Table 1e) 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.

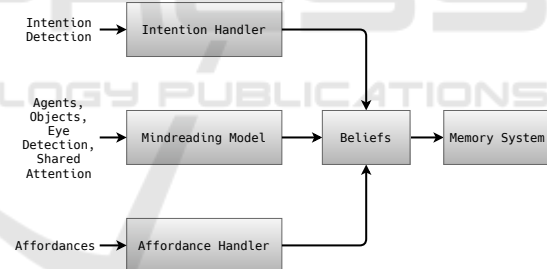


Figure 4: CogToM Processing Blocks.

CogToM Cognitive Architecture consists of four main processing blocks (Figure 4), the Mindreading Model, an Affordance Handler, an Intention Handler and a Memory System. The processing blocks create internal representations of the input data (Table 2).

The **Mindreading Model** block implements the four major modules (ID, EDD, SAM, ToMM) as defined by Baron-Cohen (Baron-Cohen, 1997).

*ID* is responsible for identifying agents and objects in the scene and takes as input Table 1a data, creating an internal representation as the agents Table 2a.

*EDD* takes as input Table 1b data and creates internal representations: an Entity Table for the agents and objects, an Eye Direction Table that reports which agents are seeing what objects (a value of *1* indicates

Table 2: Internal model tables for the canonical false-belief test.

(a) ID Agents Table		(b) EDD Entities Table			(c) EDD Eye Direction Table				
Agents		Entities			Sally	Anne	Basket	Box	Ball
Sally		Sally			Sally	0	1	1	1
Anne		Anne			Anne	1	0	1	1
		Basket							
		Box							
		Ball							

(d) EDD Agent Registry					(e) SAM Shared Attention Table		
Agent Registry					Object	Agent 1	Agent 2
Sally	Anne	Basket	Box	Ball	Basket	Sally	Anne
Anne	Sally	Basket	Box	Ball	Box	Sally	Anne
					Ball	Sally	Anne

(f) TOMM Beliefs Table					(g) Observer Beliefs Table				
Agent	Belief	Object	Affordance	Target	Agent	Belief	Entity	IS AT	Target
Sally	Believes	Anne	Exists	None	Observer	Knows	Sally	IS AT	Room
Sally	Believes	Basket	Contains	None	Observer	Knows	Anne	IS AT	Room
Sally	Believes	Box	Contains	None	Observer	Knows	Basket	IS AT	Room
Sally	Believes	Ball	Hides	None	Observer	Knows	Box	IS AT	Room
Anne	Believes	Sally	Exists	None	Observer	Knows	Ball	IS AT	Room
Anne	Believes	Basket	Contains	None					
Anne	Believes	Box	Contains	None					
Anne	Believes	Ball	Hides	None					

the object is in the visual field of the agent, 0 otherwise), and an Agent Registry that based in the Eye Direction Table creates a list of the objects in the visual field (tables 2b, 2c, 2d).

*SAM* creates three-way relationships between two agents and one object (or another agent) to assign a shared attention property to the agents (Table 2e). *SAM* takes as input the internal processing results from ID and EDD.

*ToMM* creates textual presentations in the form of *Beliefs* as a result of the visual processing in the mindreading modules (Tables 2f and 2g). There are two sets of beliefs the *Observer* generates: beliefs associated to each of agents in the scene and self-beliefs of the *Observer*, such as the location of all entities in the scene.

The **Affordance Handler** block is responsible for processing object affordances and assigning properties to the objects in a scene from Table 1c data, modifying the set of beliefs in the *ToMM* module.

The **Intention Handler** block is responsible for assigning outcomes for agent Intentions from Table 1d data and, based on the object affordances, also modifying the set of beliefs in the *ToMM* module.

The **Memory System** block stores the results of the simulation in the form of textual descriptions, assigning the *Beliefs* of each agent in each simulation step.

*CogToM* provides a console system to control processing of input for the duration of the simulation. The console stops the simulation after each mind step is processed and allows querying the mental states of the *Observer* entity.

The mental states of the observer entity are returned as *Beliefs*, text descriptions of the form:

< AGENT, BELIEVES/KNOWS, OBJECT, AFFORDANCE, TARGET OBJECT >

Where:

- **AGENT:** is the main agent that the mental state applies to, for example *Sally*.
- **BELIEVES/KNOWS:** are the mental states assigned to the agent and to the observer. There are various mental states that could be considered including the states of pretending, thinking, knowing, believing, imagining, guessing and deceiving. For the purposes of this system, only the *Believe* and *Know* mental states are initially implemented.

- **OBJECT:** is the object of the belief, for example *Ball*.
- **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*.

The console system allows for queries on the mind state of the *Observer* entity regarding the agents and objects of the scene, providing output as the set of beliefs:

```

Anne Believes Ball Hides
Anne Believes Basket Contains
Anne Believes Box Contains
Anne Believes Sally Exists

Observer Knows Anne IS AT Room
Observer Knows Ball IS AT Room
Observer Knows Basket IS AT Room
Observer Knows Box IS AT Room
Observer Knows Sally IS AT Room
    
```

## 6 RESULTS

### 6.1 The Sally-Anne Test

The canonical false-belief test is described by the sequence of steps:

#### 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.

To assess CogToM’s ability in Sally-Anne test, we simulated a visual system (Grassiotto and Costa, 2020) that would be capable of generating tables 1a, 1b and 1c (the tables show just the first two mind-steps). Also we simulated a system that is capable of, given an object, returning typical affordances of the object, as in Table 1c.

CogToM is able to reply correctly to the belief question as described in the Sally-Anne test as can

be seen in the console output below, providing the set of beliefs:

```

Anne Believes Ball OnHand Of Anne
Anne Believes Basket OnHand Of Anne
Anne Believes Box Contains
Anne Believes Sally Exists

Sally Believes Anne Exists
Sally Believes Ball HiddenIn Basket
Sally Believes Basket Contains
Sally Believes Box Contains
    
```

This can be seen in the two mental models for Anne and Sally:

```

Anne Believes Ball OnHand of Anne
Sally Believes Ball HiddenIn Basket
    
```

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 is able to pass the false-belief task.

### 6.2 The bAbI Dataset

In order to measure the performance of the system with tasks other than the canonical false-belief task, other scenarios were considered for validation.

Facebook Research proposed The bAbI dataset in (Weston et al., 2015) as a set of 20 simple toy tasks to evaluate question answering and reading comprehension. Two tasks from this set were considered for the validation of the proposed architecture.

#### Task 1: Single Supporting Fact.

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

**Task 1:** consists of a question to identify the location of an agent, given one single supporting task (*Mary travelled to the office*). Input tables for this first task are at Table 3.

In order to handle positioning questions, the system uses positioning data at Table 3b to reply correctly using the *Observer* beliefs for the location of agents and objects in the environment.

CogToM was able to respond correctly the location of the agents, by querying for the *Observer* beliefs:

```

Observer Knows John IS AT Hallway
Observer Knows Mary IS AT Office
    
```

**Task 2:** is an extension of task 1 to query the location of an object that is on hand of an agent. The *Observer*

Table 3: Input tables for Facebook bAbI Task 1.

(a) entities.txt			(b) positioning.txt		
t	Entity	Is_agent	t	Entity	Location
1	Mary	True	1	Mary	Bathroom
1	John	True	1	John	Hallway
2	Mary	True	2	Mary	Office
2	John	True	2	John	Hallway

(c) eye_directions.txt			(d) affordances.txt	
t	Agent	Object	Object	Affordance
1	Mary	John	Mary	Move
1	John	Mary	John	Move
2	Mary	John		
2	John	Mary		

(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

## 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: Input tables for Facebook bAbI Task 2.

(a) positioning.txt		
t	Entity	Location
1	John	Playground
1	Bob	Playground
1	Football	Playground
2	John	Playground
2	Bob	Kitchen
2	Football	Playground

beliefs support the positioning of agents and objects as can be seen on Table 4. For brevity, only the positioning table is shown here.

The console output shows the *Observer* beliefs with the location of the football.

```
John Believes Bob Exists
John Believes Football OnHand Of John
```

```
Bob Believes Football OnHand Of John
Bob Believes John Exists
```

```
Observer Knows Bob IS AT Kitchen
Observer Knows Football IS AT Playground
Observer Knows John IS AT Playground
```

## 7 CONCLUSION

CogToM was designed as a platform to validate the viability for a computational system to pass false-belief tasks based on implementing a psychological model of the human mind. The system thus designed has shown us the need for integrating further information about the world in the form of affordances and human intentions.

The cognitive architecture we proposed is 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 simple tasks as described by the Facebook bAbI dataset.

Even though the main motivation for the design of CogToM was to explore the viability of designing an assistant for people in the autism spectrum, our system 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.

CogToM holds promise as a base system from which future assistive systems for people in the autism spectrum can be based on.

## ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## REFERENCES

- Augustyn, M. (2019). Autism spectrum disorder: terminology, epidemiology, and pathogenesis.
- Baraka, M., El-Dessouky, H. M., Abd El-Wahed, E. E., Amer, S. S. A., et al. (2019). Theory of mind: its development and its relation to communication disorders: a systematic review. *Menoufia Medical Journal*, 32(1):25.
- Baron-Cohen, S. (1990). Autism: A specific cognitive disorder of "Mind-Blindness". *International Review of Psychiatry*, 2(1):81-90.
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

- Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme*, 34(1):74–183.
- Baron-Cohen, S., Leslie, A. M., Frith, U., et al. (1985). Does the autistic child have a “theory of mind”. *Cognition*, 21(1):37–46.
- Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., and Chetouani, M. (2014). Interactive technologies for autistic children: A review. *Cognitive Computation*, 6(4):722–740.
- Grassiotto, F. and Costa, P. (2020). CogtoM simulation results. Last checked on Sep 08, 2020.
- Jaliaawala, M. S. and Khan, R. A. (2020). Can autism be catered with artificial intelligence-assisted intervention technology? a comprehensive survey. *Artificial Intelligence Review*, 53(2):1039–1069.
- Kientz, J. A., Hayes, G. R., Goodwin, M. S., Gelsomini, M., and Abowd, G. D. (2019). Interactive technologies and autism. *Synthesis Lectures on Assistive, Rehabilitative, and Health-Preserving Technologies*, 9(1):i–229.
- Kimhi, Y. (2014). Theory of mind abilities and deficits in autism spectrum disorders. *Topics in Language Disorders*, 34(4):329–343.
- Klin, A. (2006). Autism and asperger syndrome: an overview. *Brazilian Journal of Psychiatry*, 28:s3–s11.
- McClelland, T. (2017). Ai and affordances for mental action. *environment*, 1:127.
- Montesano, L., Lopes, M., Bernardino, A., and Santos-Victor, J. (2008). Learning object affordances: from sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, 24(1):15–26.
- Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3575–3584.
- Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526.
- Şahin, E., Çakmak, M., Doğar, M. R., Uğur, E., and Üçoluk, G. (2007). To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4):447–472.
- Scassellati, B. M. (2001). *Foundations for a Theory of Mind for a Humanoid Robot*. PhD thesis, Massachusetts Institute of Technology.
- Weston, J., Bordes, A., Chopra, S., Rush, A. M., van Merriënboer, B., Joulín, A., and Mikolov, T. (2015). Towards ai-complete question answering: A set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698*.
- WHO (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*, volume 2. World Health Organization.
- Yi\*, K., Gan\*, C., Li, Y., Kohli, P., Wu, J., Torralba, A., and Tenenbaum, J. B. (2020). Clevrer: Collision events for video representation and reasoning. In *International Conference on Learning Representations*.
- Yu, Z., Kim, S., Mallipeddi, R., and Lee, M. (2015). Human intention understanding based on object affordance and action classification. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE.