BELIEF REPRESENTATION SYSTEMATIC APPROACH (BRSA):

AN AGENT-BASED MODEL TO UNDERSTAND

A SIMPLE THEORY OF MIND

by

ZAHRIEH YOUSEFI

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

August 2016

ABSTRACT

A meaningful social life relies on understanding others' minds and behaviours. Theory of mind (ToM) is the ability to reason about an individual's mental states such as beliefs and desires, and to understand and predict how these mental states shape an individual's behaviour.

This thesis aims to develop a systematic approach for understanding the underlying processes of a simple theory of mind and to evaluate the performance of theory of mind ability in a social context. For this purpose, two case studies using agent-based modelling methodology has been conducted.

An original set of basic processes underpinning ToM ability, termed Belief Representation Systematic Approach (BRSA) has been explored through these two models. BRSA reconstructs ToM processes into four main phases: Perception, Memory, Reasoning beliefs and desires, and Action.

BRSA clarifies that there is a difference between having ToM and 'using' it. The reasoning involved in the third and fourth phases of BRSA influences the agents' performances. BRSA shows that false belief tasks require two preconditions, resources and reasoning, to be considered as an acid test for ToM competence.

Both models demonstrate that developing agents' understanding of others' mental states on the micro level will lead to significant improvements in their social performances on the macro level.

To Sodabeh and Gharib, my late parents

for their incredible love

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my PhD supervisors, Dr Dietmar Heinke and Professor Ian Apperly, for their invaluable guidance. Thank you for your scientific instructions and insightful comments on this project.

Dietmar, thank you for your constant advice and patient supervision over these years - your thoughtful opinion has been a vital assistance to my work. I am particularly grateful to you for ensuring the results were rigorously evaluated.

Ian, your extraordinary expertise on theory of mind was an inspiration to me throughout the project, and without your academic guidance and encouragement, it would not have been possible to conduct this research.

I am very grateful to Dr Sarah Beck for her administrative support throughout the project.

In addition, I would like to thank Professor Aaron Sloman for spending his valuable time with me to share his wonderful ideas.

My special thanks to all my dear friends, Kate, Janet, Afsane and Jacki for their uplifting and reassuring conversations. In particular, I appreciate Frances's constant inspiration and advice and Sandy's support.

Thank you to my caring brothers and sisters; Pari, Noroz, Yalda and Shahin, for their love and encouragement from the other side of the ocean.

Last but not least, many thanks to my lovely son, Navid for his understanding and patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1. GENERAL INTRODUCTION

"Everyone became my friend from his own perspective; none sought out my hidden feelings from within me."

Maulana Jalaluddin Rumi

"[Common sense is] the mental skills that most people share. Common sense thinking is actually more complex than many of the intellectual accomplishments that attract more attention and respect, because the mental skills we call "expertise" often engage larger amounts of knowledge but usually employ only a few types of representations. In contrast, common sense involves many difficult kinds of representations and thus requires a larger range of different skills."

(Marvin Minsky, 1988, pp. 327)

1.1 INTRODUCTION

Humans are a social species; they understand their own and others' behaviour in their day-to-day life. Their capability to make spontaneous inferences about others' invisible thoughts and feelings enables them to communicate with others.

Humans' natural ability to mentally process others' behaviour is central to their social life. This mental process relies largely on understanding the essential constituents of others' behaviour such as their beliefs, desires and goals. Unsurprisingly, people infer others' mental states in everyday life; adults are competent in flexible and complex social reasoning while infants start from tracking eye direction effortlessly and automatically (Apperly & Butterfill, 2009). Intriguingly, from the simple to the complex, mental processes influence the actions people take to reach their goals. How do humans infer others' mental states? How does understanding others' beliefs and desires improve one's

performance in terms of achieving her goals? What are the basic processes through which we can understand others' mental states? What are the heuristics employed in this process? A large body of research addressing such questions consistently suggests that humans have theory of mind; the ability to take others' mental states into account and apply this coherent information to infer their actions (Frith, 2012). People understand, predict and even defend others' actions by reasoning about others' beliefs and desires. They distinguish others' belief, desires and goals from their own one, they use others' beliefs and desires to rationally predict their actions. Moreover, they may reason backwards to infer others' beliefs and desires from their actions (Baker, Saxe, & Tenenbaum, 2009).

The nature of theory of mind has become more clearly understood through ever increasing number of studies in the last three decades (Wellman, 2014). Yet, the areas of confusion and gaps are increasing and lack of standardization in the literature has been recently identified by researchers (e.g. Apperly, 2012; Schaafsma et al., 2015). For example, some of the experimental methods involved abilities, which are not related to theory of mind and have therefore misinterpreted the theory of mind processes.

The problem arises because some researchers tacitly regard theory of mind as a single, indivisible process, some consider it as a single brain network and other researchers combine varieties of theory of mind into one process (Schaafsma et al., 2015). Despite an increasing number of studies, theory of mind literature lacks a systematic approach to its basic processes, leading to confusion in many of its experiments and the results that arise from them.

The objective of this thesis is to develop a systematic approach for understanding the principles of a simple theory of mind, thinking about others' beliefs, desires and goals, and to evaluate the performance of theory of mind ability in a social context. For the purpose of this thesis, two case studies using agent-based modelling have been conducted. Agent-

based models comprises of agents interacting in an environment. Typically, agent-based models facilitate the simulation of social processes through the interactions between individual agents (micro level) that can generate certain social phenomena (macro level). Agent-based models are capable of representing the rules and relationship between individual's mental states, their actions, the environment and the way that they infer others' mental states on the micro level. Whilst the macro level represents the effect of these actions, for example, how successful individuals perform in the environment. These characteristics of agent-based modelling provide a reliable framework for interactions between the individual's mental states and their actions with others within the virtual society.

This thesis argues for three theoretical and one methodological points. Firstly, this study offers a robust and transparent systematic set of basic processes comprising a simple theory of mind ability including the ability that enables success on some key tasks, such as false belief tasks. This systematic approach consists of four main phases.

Secondly, the systematic approach proposed by this thesis demonstrates that false belief tasks, as a common decisive methodology for theory of mind competence, might involve more than understanding others' beliefs.

Thirdly, reasoning is a central information processing part of theory of mind. The various levels of theory of mind originate from different levels of reasoning. The lowest level of reasoning may produce an automatic version of theory of mind.

Fourthly, the agents, which are able to construct inferences about others' beliefs, desires and goals, perform more effectively than agents that only recognise their own beliefs and desires.

To the best of the author's knowledge, this thesis is the first attempt to explain underlying processes of a simple theory of mind through an agent-based model and the first model to evaluate the link between the potential performances of agents with their simple level of theory of mind ability in a competitive society.

1.1.2 Thesis Structure

The first chapter of this thesis provides a general introduction to the two subjects of this study; theory of mind and agent-based modelling. This chapter summaries the core terminology and background literature in these two realms.

The empirical chapters consist of an introduction, a detailed description of the methodology, hypotheses, predictions and implementation of the model, varieties of simulation runs, the results, a discussion regarding the simulation results and a conclusion in addition to the related questions in the field and a comparison between the simulation results and the developmental literature.

The first empirical chapter, chapter 2, represents an agent-based model to understand the abilities that might underlie success on standard false belief tasks. Three types of agents' strategies and their diagrams are depicted, along with the explanations of the impact of their different theory of mind abilities in their performances, the cognitive phases that occur within the agents, and the resources they require for their abilities.

The second set of simulations is provided in chapter 3. This model consists of a variety of agents with different levels of understanding of their own and others' mental states, clarifying how the agents' micro level rules have impact on their macro level behaviours. The agents' range of abilities varies from having no understanding of own and others' mental states to being able to infer others' mental states through their actions. The relationship between agents' different abilities and their performances is discussed in depth. Chapter 3 describes the advantages of understanding others' beliefs and desires

though their actions in a competitive environment. The study outlined in chapter 3 identifies a standard approach to theory of mind processes similar to the first model in the previous chapter. In light of this approach, chapter 3 clarifies the main distinction between minimal theory of mind ability and theory of mind competence in the simulation. The outcome of the simulation is compared with the relevant theories in developmental psychology literature.

Chapter 4 specifically focuses on the main shared result of both simulations, a novel standard approach for belief representation processes. The aim of this chapter, therefore, is to develop on this concept, explain where it comes from, and what its implications in the literature are. This important chapter illuminates how this standard approach can be beneficial in explaining some of the controversial issues in the literature. The standard approach is discussed in depth, firstly in regards to false belief task, and secondly in terms of developmental literature. The reasons of failure in false belief tasks, the link of this approach with minimal theory of mind and the conditions in which the false belief task is a decisive test for theory of mind are explained in this chapter.

The final chapter of this thesis provides an overview of the main results. In chapter 5, an endeavour is made to shape these findings into a general framework that addresses theory of mind processes in a systematic way. Moreover, the necessary resources needed to enable agents to infer others' mental states, and the role of each of these resources are explained. The discussion regarding the important role of reasoning in belief representation, and also the effect of theory of mind ability on agents' performances is presented in this chapter, through interpreting the results of the models. Finally, the limitations and possible future work of both models are explained at the end of this chapter.

Key Words

False Belief task, Theory of mind, Minimal theory of mind, Belief Representation, Agent-Based model.

1.1.3 Glossary

Inhibition (cognitive inhibition):

"Cognitive inhibition is the stopping or overriding of a mental process, in whole or in part, with or without intention. The mental process so influenced might be selective attention or memory retrieval or a host of other cognitive processes.*"* (MacLeod, 2007, p. 3)

Self-perspective inhibition:

The ability to temporally prevent one's own perspective to consider others' perspective (Samson et al., 2005).

Belief-like:

A simple form of mental content from recent visual experience in which the information is no longer immediately available but registered (Apperly & Butterfill, 2009).

Mental states:

The condition of activity of any simple part or process of the mind at a certain point in time (Minsky, 1988).

1.2 THEORY OF MIND (ToM)

How could humans be able to participate in their social life without understanding others' beliefs, intentions, desires or emotions? Imagine humans could not attribute or interpret others' mental states at all or inversely everybody could completely understand others' beliefs, intentions, desires and feelings in human society. Both cases might indirectly result in a perfect chaos or perfect mindreading. However, this is not the case - the real human society differs entirely from both situations where perfectly understanding others' mental state or completely not making sense of others.

Since the mental states of others cannot be directly perceived, it is commonly supposed that theory of mind involves one or more processes of inference, whereby representations of others' mental states are inferred from what they do or say.

Premack and Woodruff were the first to coin the term "Theory of Mind" in 1978 by asking a crucial question in an article: "Does the chimpanzee have a theory of mind?" A system of inferences about others' mental states that are not directly observed can be used as a theory to make predictions about others' behaviour. They define that one has a theory of mind if one could impute mental states to self and others (Premack & Woodruff, 1978).

"Mindreading" is another theoretically natural term that captures the characteristics of the problem more than 'theory of mind' (Apperly, 2011), and allows us to understand and explain meaningful behaviour and actions of other people and predict or influence individual's behaviour.

After more than thirty years of extensive research in theory of mind, Apperly (2012) argues that 'consensus' on what theory of mind is and how it could be studied, requires revision. He suggests three distinct approaches to study theory of mind: The conceptual domain, the cognitive perspective and the social competence that can vary across individuals. Apperly suggests that future research will benefit from clearly expressing in what aspect of theory of mind we wish to measure, because each of these directions considers different questions that require different approaches. A concise literature review based on the conceptual domain and the cognitive domain is described below, after the explanation of false belief task, two system account and minimal theory of mind.

1.2.1 False Belief Task

From its inception, theory of mind research has evolved with various experiments to assess different characteristics of theory of mind abilities in non-human and human children. However, one type of tasks, false belief tasks have been the most frequently used task and

have stimulated the developmental research widely (e.g. Apperly, 2011; Doherty, 2009; Bloom & German, 2000). There are various styles of false belief tasks, but all methodologically are based on a perspective difference between the child and the target by making a child friendly story (Apperly, 2011). The most common and standard task relies on a change of the location of an object.

The logic behind false belief task was first outlined by (Dennett, 1978) , in a review on chimpanzees' theory of mind by Premack & Woodruff (1978). Dennett reasoned that to avoid the misunderstanding between associated behaviour and mental state attribution, it is essential to clearly test the chimpanzees' mental state attribution by giving the subject a false belief. This was in response to the general doubt that some of the experimental set ups that involve only some physical and behavioural events rather than unobservable mental states to examine theory of mind ability. Wimmer and Perner (1983) were the first that implemented Dennett's idea to an "unexpected transfer task" as the first false belief task for children (Hedger & Fabricius, 2011).

1.2.2 Verbal false belief task (Explicit)

Baron-Cohen et al. (1997) revised Wimmer and Perner's task to a simpler version for children to understand. The task, so called "Sally and Ann false belief task", has been used widely in the literature and is identified as standard false belief task to test children's ability in understanding others' beliefs.

The scenario of the task includes two puppets, Ann and Sally. The subject child who is being tested for belief representation watches as Sally puts the ball in the basket. Then Sally leaves the room and when Sally is out, Ann moves the ball from the basket into the box. Sally returns, and the child is asked where Sally will look for the ball. The Sally and Ann false belief task is illustrated in Figure 1.

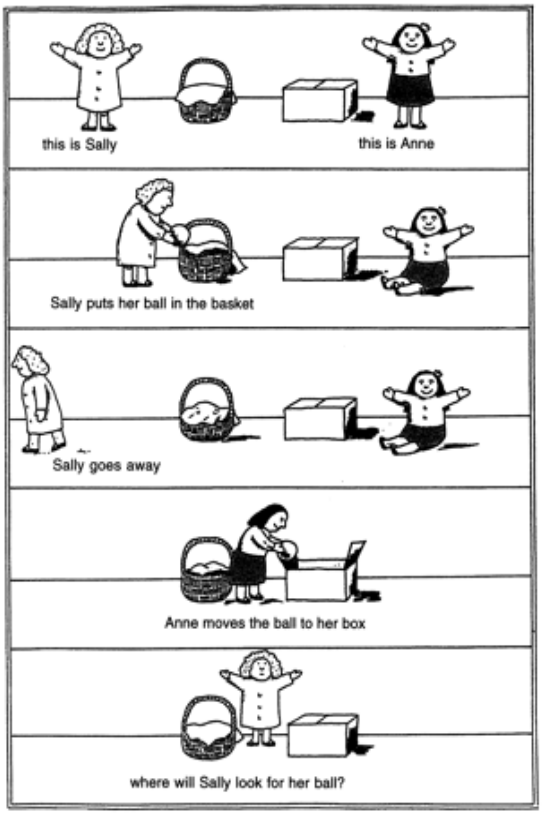


Figure 1. Sally and Ann False Belief Task from Baron-Cohen et al. (1997)

The general design of false belief tasks solves the behavioural aspect of understanding others' beliefs by introducing two separate beliefs about the location of an object (the ball), one is the real location of the object and another is the protagonist's (Sally's) perspective that is a false belief about the location of the object. In standard false belief task, the real location of the ball is in the box, where Ann moved the ball. However, in Sally's perspective, the location of the ball is in the basket where she last put and saw the ball. The child needs to predict where Sally will look for the ball when she returns. The correct answer to this question is the basket where Sally put the ball. The child needs to answer the question correctly to pass the false belief test.

The standard false belief task is a verbal explicit task and the child is required to reason about the Sally and Ann scenario to answer a question correctly. The child needs to distinguish Sally's belief about the location of the ball from his/her own perspective that is

different (because the child saw Ann moved the ball and Sally did not see). If the child passes the standard false belief task, then he/she is thought to have the conceptual competence to understand others' beliefs.

The Sally and Ann false belief task is a change of location task. There is an alternative false belief task, called the "unexpected content" task. For example, the subject child sees a smarties box. The child then opens the box and sees there is a pencil in the box instead of smarties. The verbal reasoning question for the child is that what someone who has never observed inside the box will think the box holds inside. The correct answer is smarties, which is different from the child's perspective (as the child has observed that there is a pencil inside the box whereas the participant will have not).

The false belief task is often considered as an acid test for theory of mind because it shows the understanding of others' different perspective. The results of the both verbal false belief experiments, the Sally and Ann task and the unexpected content task, show that children under the age of 4 years fail the verbal false belief task (e.g. Wellman, Cross, & Watson, 2001; Call & Tomasello, 2008; Wellman, 2014).

There is also evidence that if the false belief task is simplified, most children at age 3 are able to pass the false belief task. For example, the question of the task was more specific and easier for children to understand, by assisting the child to remember the content of the box or by reducing the salience of the child's own perspective (e.g. Lewis & Osborne, 1990; Siegal & Beattie, 1991; Freeman, Lewis, & Doherty, 1991; Mitchell & Lacohee, 1991; Moses, 1993; Freeman & Lacohee, 1995; Carlson, Moses, & Hix, 1998; Wellman & Bartsch, 1988; Surian & Leslie, 1999; Yazdi et al., 2006). Moreover, these experiments indicate that failure to pass the standard false belief task is due to children's inefficient processing capabilities (German & Leslie, 2000).

However, the meta-analysis by Wellman (2014) showed that the beneficial effects observed in these studies were small and children under 4 years are not capable of representing beliefs. Wellman describes two different meta-analyses including 600 variations of children's false-belief performance as follows:

The first meta-analysis comprising 178 studies with various false belief task manipulations, making it easier for children to understand (Wellman, Cross, & Watson, 2001). The second one includes experiments of false belief tasks across different cultures and languages (Liu et al., 2008). The results reveal that children in preschool years, at around the age of four, develop theory of mind universally.

Although Wellman's meta-analysis produces findings compatible with conceptual change, it does not determine any nature of mechanisms of theory of mind theoretically (Scholl & Leslie, 2001).

1.2.3 Non-verbal false belief task (Implicit)

Verbal (explicit) false-belief tasks are naturally complex, as they require the integration of linguistic information. In addition, these tasks may generate disruption for children in the process of tracking events from the protagonist's point of view (Rubio-Fernández, 2013). Thus, researchers start to design nonverbal (implicit) false belief tasks with less cognitive demands to test infants' belief representation competence.

Onishi and Baillargeon (2005) developed a non-verbal false belief task for 15-month-old infants concerning the change of an object location by using the "violation of expectation" (VOE) paradigm and measuremet of their looking time. Infants see that the protagonist places the object in a box (black box). Then protagonist leaves and does not see the change of the object's location to the white box. When the protagonist returns to the scene the infant expects her to be searching for the object. The results confirm that infants look longer at the white box. Therefore, by applying the VOE paradigm, infants

expect the protagonist to search the black box for the object as the infant looks longer at the white box which violates her/his expectation. The findings from this experiment and others (e.g. Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007; Baillargeon, Scott, & He, 2010; Kovács, Téglás, & Endress, 2010; Southgate & Vernetti, 2014) indicate that infants are able to pass non-verbal false belief tasks as young as 7 months, and in any case well below 4 years old. This contradiction has been a pivotal debate and issue in developmental literature, producing fruitful research and raising more questions.

1.2.4 Discussion of verbal versus non-verbal false belief tasks

The questions raised regarding the contradiction between verbal and non-verbal false belief tasks. Why preschool children fail in verbal false belief task whereas infants are able to pass non-verbal false belief tasks? Does it demonstrate infants' attribution of understanding others' beliefs? What is the difference between understanding of verbal false belief task versus non-verbal one? If the methodology is correct, how should we interpret the contradiction?

Various valuable approaches explored their findings and suggested their perspectives to these controversial issues. All of these different opinions agree on the experimental results, and also that preschool children fail the explicit false belief tasks. Intriguingly, however, there is no consensus on whether infants understand implicit false beliefs. Although, there is a lack of consensus, these plausible opinions would provide fresh insights in understanding belief representation. Broadly speaking, there are three different explanations for this problem. The first explanation by Wellman suggests this is due to the developmental changes in preschool children. He argues that implicit false belief tasks require a simple attribution of desires and awareness rather than beliefs and false beliefs (Wellman, 2014). In general, he considers theory of mind as set of concepts which need to be acquired.

The second account argues that under 4 years children's belief representation abilitiy is obscured through the complexities and tasks demands; for example the child participants in false belief task need to override their own natural belief about the right location of the object and point out other's belief which is not the right location. This may be a prepotent response that a child struggles to inhibit ("inhibition of prepotent responces"). According to the researchers of this line of argument (e.g. Leslie , 2005; Southgate & Vernetti, 2014; Baillargeon, Scott, & He, 2010; Kovács, Téglás, & Endress, 2010) infants are basically competent at belief-desire reasoning likewise preschool children, and just fail to show this competence on tasks that exceed their capacity to inhibit inappropriate responses.

The third account is offered by Apperly and Butterfill suggesting an implicit system for infants which is fast and limited, it enables infants to track others' belief but not psychological representation as such whereas preschoolers' belief undersanding is explicit, flexible and cognitively demanding (Apperly & Butterfill, 2009). They suggest that infants may have an automatic emerging system that does not need much experience and it persists into adulthood (Apperly & Butterfill, 2009).These two systems will be further described later in this chapter.

1.2.5 Two Systems Account

Apperly and Butterfill have advocated 'two systems' account, as a parallel cognitive construction of theory of mind. The nature of these two distinct systems rests on a compromise between flexibility and efficiency. The first system is fast and cognitively efficient and capable of tracking others' registration of an object rather than belief representation as such but inflexible and limited. They suggest that such a system may account for the success on some theory of mind tasks by human infants, some non-human animals such as chimpanzees and human adults under cognitive load. Whereas the second system associates with a cognitively demanding but flexible and slow processing. The

second system exists in human adults, parallel with the first system. They claim that there is only limited information flow between two systems. In addition, system one is automatic whereas system two is non-automatic and requires reasoning.

They argue that the two systems account is a promising solution to a major puzzling pattern in theory of mind literature; the contradiction that infants pass the nonverbal false-belief task, yet children pass the verbal false belief task at around 4 years of age. Moreover, it answers to the polarized point of human adult efficiency and quick responses in their social interactions that demands theory of mind (Apperly & Butterfill, 2009).

1.2.6 Minimal Theory of Mind

Butterfill and Apperly (2013) proposed a novel approach to minimal theory of mind to represent the ability to track others' perception, knowledge and beliefs in primitive species, infants with limited cognitive resources and human adults under load. In fact, this is an elaborated version of the first system of their two systems account. Apperly and Butterfill developed a distinctive minimal form of theory of mind cognition that purely involves representing "belief-like" states without any cognitive demands or conceptual sophistication. Their argument starts with a fundamental question that "what could someone represent that would enable her to track, at least with limits, other's perception, knowledge states and beliefs including false beliefs?" (Butterfill & Apperly, 2013, p. 1) They then have formulated four principles to answer this question.

The first principle relates to a basic concept of goal-directed action. "A minimal grasp of goal-directed action" (Butterfill & Apperly, 2013, p. 10) that one might understand the goal from bodily movements such as tracking others' visual direction or a change in gaze direction.

In the second principle, Butterfill and Apperly (2013) introduce two terms, "field" and "encountering" as a basic characteristics of perception. The concept of field relates to a set

of objects at any given time that is determined by factors such as proximity, eye direction and barriers. The concept of encountering is a relation between the agent and an object that is a true concept only if the object is in the agent's field. According to Apperly and Butterfill, to act goal-directly on an object, it is necessary to encounter it first.

The third principle by Butterfill and Apperly is called "registration", which introduces a new notion of a belief-like concept. Registration is a relation between an agent, an object and its location. The agent registers the location of an object as it encounters the object. A correct registration is a precondition for a successful goal-directed action. Apperly and Butterfill state one's correct registration of an object becomes incorrect by moving or destroying the object in her/his absence.

One example for an application of the registration principle is scrub-jays re-cache food experiment by Clayton et al. (2007). In this experiment, scrub-jays have only chosen to re-cache the food in the presence of competitors who previously saw they cached it. Apperly and Butterfill suggest that scrub-jays understand that competitors' correct registration of food results to stealing their cached food as a successful goal-directed action. Therefore, they re-cache the food to prevent competitors from correctly registering its location (Butterfill & Apperly, 2013).

Another example of the third principle is an experiment by Liszkowski et al. (2006). In this experiment, infants are pointing towards the locations of the accidentally misplaced objects to provide relevant information to adults. Apperly and Butterfill suggest that pointing expresses a correct registration and infants understand that correct registration results to a successful goal-directed action.

The fourth principle involves a shift to thinking of a successful registration as a causal factor for the agent's action. In other words, "when an agent performs a goal-directed action

with a goal that specifies a particular object, the agent will act as if the object were in the location she registers it in." (Butterfill & Apperly, 2013, p. 16)

The nonverbal false belief task by Onishi and Baillargeon (2005) is an application of the fourth principle. Infant subjects and an observer watch while an object is placed in a black box. In the absence of observer, the object is moved to a white box. When the observer comes back, infants looking times indicate that they correctly expect that the observer will reach into the black box. Onishi and Baillargeon suggest that the infants are ascribing beliefs about the object to the observer. However, Apperly and Butterfill's alternative explanation is that infants track the registered location of the black box as a cause of action. They present an insightful example of minimal theory of mind. Suppose Hannah is able to distinguish whether someone can see her while she is stealing from others. She wants to escape others' detection by "exploiting a fact about other's mental states (namely that they usually cannot see Hannah's acts of theft when Hannah does not have their eyes in view). Then Hannah has a theory of mind ability" (Butterfill & Apperly, 2013, p. 606). They continue that Hannah is able to use others' visual perspective without any theory of mind that requires complicated cognitive ability. More importantly, they suggest that an individual with minimal theory of mind ability could pass many tests that were supposed to be acid tests of theory of mind such as false belief tasks. The reason for this is that minimal theory of mind does not require sophisticated resources for reasoning.

Recently, the literature on theory on mind has increased greatly with a variety of methodologies and experiments. Thus, the aim of the following sections is to briefly review the literature related to the models proposed in this thesis, the two domains, the conceptual domain and the cognitive domain.

1.2.7 The Cognitive Perspective

The cognitive perspective involves the architecture and process of theory of mind; it clarifies the way that belief representation works at a cognitive level. For example, the role of language in false belief task process and the way that people inhibit their own perspective and take others perspective in to the account (Apperly, 2012).

Experimental research in cognitive processes of theory of mind has widely expanded for example by making inferences about others' beliefs, storing information about others' perspective in mind, and applying theory of mind in social context (Apperly, 2012).

The evidence shows that language, memory, executive functions are critical in belief representation (e.g. Apperly et al., 2007; Hughes, 1998; Marcovitch et al., 2015). The parallel link between executive function and theory of mind competence has been studied extensively, particularly the inhibitory control role in taking others' perspective into consideration and preventing egocentric version of information in children and adults. For example, Russell (1996) argues that deficit in executive control in autistic children underpin the emergence and expression of their theory of mind ability. However, Wellman (2001) states that theory of mind development corresponds with executive function but not directly.

The study by Leslie and Polizzi (1998) offers a model to pass a false belief task; the false belief reasoning starts with identifying a true belief content. They suggest that theory of mind mechanism (ToMM) nominates a true-belief to the content of the belief and it is selected to attribute as the belief state. Hence, to pass the false belief task, it is necessary to inhibit the default content of the belief, which is true, and change the attention to alternative belief. In case of inhibition failure, the default content will be allocated to the belief state, which is inaccurate in the case where the target has a false belief. In total, the sequence of false belief reasoning starts with allocating a true belief to the belief content

as a default and then inhibition starts to prevent the default and it redirects toward the alternative belief, which is a false belief.

Leslie and Polizzi (2008) attempted to expand their theory by a change in Sally's implicit desire. In Sally Ann false belief task, the desire of Sally is to find the ball. Suppose Sally has no desire to find the ball. Thus, this avoidance of the ball needs to be consider as well as the belief about the location where she put the ball. Leslie and Polizzi suggest that firstly one needs to recognise the location of the ball, and then consider the desire of avoiding the ball. Therefore, the child participant needs to prevent the default desire, and choose the location, which does not contain the ball. They concluded that this task requires two inhibitions, one for belief and one for desire, which cancel one another out.

The concept of selection processing (SP) is "to select the most plausible belief content from a small set of plausible candidates" (Leslie, German, & Polizzi, 2005, p. 51). Leslie et al. (2005) explain SP as an automatic process, which is associated with ToMM and attributes beliefs and desires to the agent. The concept of inhibition and ToMM is identified through the algorithms of the two models in this thesis and will be discussed in section 2.5.3.3.

Another aspect in theory of mind relates to the level of complexity involved in theory of mind. For example, there is a distinction between higher-level and lower-level processes of theory of mind in terms of the levels of inferences, storing and using information involved (Apperly, 2011). The question is how and why humans are capable of higher-level theory of mind. Verbrugge (2009) suggests that competition, cooperation or mixed-motive interactions may have played a role. Reasoning is a part of higher-level theory of minds process. Intriguingly, Perner and Leekam (2008) analysed the results of 12 studies of false belief task with a similar task using a photograph of the location of the object as a non-mental representation version. Their study shows that the children still cannot pass the non-mental version of false belief task, suggesting that representational demands are

not specific to the false belief understanding but they are general to non-mental representation as well. Therfore, they argue that belief representation is involved with reasoning of non-mental representation, which highlights the reasoning role in theory of mind ability.

1.2.8 The Conceptual Domain

The conceptual approach clarifies questions such as whether a child has belief representation or at what age this concept emerges or how children acquire theory of mind concepts (Apperly, 2012). The conceptual domain also includes conceptual knowledge of others' beliefs and desires that interconnects with the behaviour (Apperly, 2012). Research in human infants and children through verbal and non-verbal false belief tasks, which have been already described, indicates an effective approach in the conceptual domain of theory of mind.

In terms of adults' theory of mind, the fact that they have the capacity to perform theory of mind does not mean that they use it automatically or use it without simple errors. They are prone to egocentrism and self-perspective resistance (Apperly, 2012). Regular adults' theory of mind rests on evidence, as they are simultaneously under construction and are prone to constructive errors (Wellman, 2014). Altogether, the way that human adults use theory of mind in everyday life to infer others' thoughts is still an unanswered question.

In terms of non-human animals, there has been some controversy over theory of mind ability in non-human primates and Premack and Woodruff's question in 1978 of whether the chimpanzee has a theory of mind. This question is still open to debate and under examination. Researchers Call & Tomasello (2008) proposed that chimpanzees' tactical deception requires more than just an understanding of surface-level behaviour. They conducted several different experimental paradigms leading to the conclusion that chimpanzees understand that others see, hear and know things. Their answer to Premack

and Woodruff's question was that chimpanzees do have a theory of mind but they do not understand others in exactly the same way as the belief–desire concepts of humans. (Call & Tomasello, 2008) They also emphasize that chimpanzees do not understand false beliefs. Considering thinking as going beyond the perceived information to make inferences, they have argued that not only thinking is not exclusive in humans, but also making inferences is not either (Schmelz, Call, & Tomasello, 2011).

However, recent research that measured the relationship between body orientation and eye gaze shows some of non-human animals' understanding is limited to visual perspective taking. Research by Hare et al. (2006) explains that chimpanzees recognise what others are able to see (Level 1 perspective taking). Nevertheless, they do not understand how others see things (Level 2 perspective taking); "understand not only what is visible from a certain point of view but also how a given object is seen or presented" (Moll & Meltzoff, 2011, p. 662) because they are attributing their own preference to others (Karg et al., 2016).

Karg et al. (2014) in their experiment present a large and a small bread stick to great apes and then blocked the scene to appear reversed. Their results show that apes are able to choose based on the real size of the stick, not based on currently perceived ones. Apes did not choose similarly in a control condition which they have no previous experience of the true size of the sticks. Although chimpanzees are able to understand the difference between their own perspectives from reality, they are not able to understand that others' perspective can be false and there is not enough evidence that they are capable to deceive and create a false belief in others (Karg et al., 2016).

In the study by Martin and Santos (2014), the participants, rhesus macaque monkeys, saw the scenes in which a human agent was watching an apple moving between two boxes. They provided different scenarios of true and false beliefs, about the final location of the apple, for both the monkeys and the human agent by occluding parts of the apple's

movement from either the monkey or the agent. The results show that monkeys looked longer at scenarios that are not consistent with their own beliefs without considering other agent's beliefs. Their findings suggest that monkeys fail to represent others' beliefs whereas human infants pass the experiment test and demonstrate belief representation (Martin & Santos, 2014). Martin and Santos (2016) in their recent paper argue that primates' belief representation is limited to the relations between agents and information that is true and they are unable to represent relations between agents and untrue information. Consequently, they suggest that belief representation may be unique to humans as part of their core knowledge systems with automatic process that enable human infants to make sense of their physical and social environments (Martin & Santos, 2014).

The models proposed in this thesis are inspired from experiments with non-human animals and human infants, from minimal theory of mind account and standard false belief task that have attempted to explain how some simple forms of theory of mind may be possible.

This thesis mainly concentrates on the cognitive perspective by representing a systematic approach for belief representation and testing agents' efficiency in a virtual society. It also explores the concept of a simple theory of mind in a social environment that is interconnected with ToM processes and reasoning at the cognitive level. In addition, this thesis examines the conceptual domain by considering which agents are capable of various levels of simple theory of mind ability (including minimal ToM, belief representation and inferring others' mental states). In other words, this thesis examines conceptually which agents are capable of theory of mind ability and why.

1.3 AGENT-BASED MODEL (ABM)

Over the years, agent-based models have been extensively emerged in various research fields, particularly in computer science, economics, sociology, psychology, philosophy and

cognitive sciences. Particularly, computational cognitive models in ABMs have increasingly combined into social and behaviour sciences (Schunn & Gray, 2002).

An agent-based model (ABM) consists of independently operating agents that are able to perceive, make decisions and perform goal-directed actions (Yilmaz, 2015). They are also known in literature as multi-agent systems.

ABMs enable us to simulate a real phenomenon to an artificial society. The idea of the simulation refers to designing a computer program model that produces the key features of the real phenomenon. By rerunning this program with various parameters, it would be possible to analyse and understand or predict the behaviour of the phenomenon. This way, agent-based models provide a generative empirical research approach to sciences. They offer a natural environment to tackle interdisciplinary study problems. Applying agent-based modelling for complex systems makes it possible to trace the problems analytically and computationally. ABMs are composed of agents, environments and interactions (Wilensky & Rand, 2015), and may results to emergence that are the main topics of this section.

1.3.1 Agent

Agents are the basic and active components of agent-based modelling. Agents are defined by their properties and their actions. Agents' properties include their internal and external states such as the agents' shape, colour, size, location, speed and direction. Agents' actions are a set of rules that agent can use to govern its behaviour and make change to the environment, other agents or itself (Wilensky & Rand, 2015). Agents are capable of autonomously making decisions to interact with each other in a virtual environment that builds up the system behaviour. Casti (1997) suggests that agents should have a higher level of actions to be able to change their own behaviour.

1.3.2 Environment (Space)

Humans and animals interact and communicate with each other in an environment. Similar to the natural world, agent-based models designs also consist of such an environment that makes interactions between the agents possible.

The agents are placed in an environment, which might consist of inactive objects such as obstacle, energy resource, roads or food. Agents interact with each other through the environment. The environment also stores the positions of the agents (Gilbert, 2007). Spatial models, in which the environment represents a geographical space, have coordinates to indicate their location like grids. Network models are designed to link agents together with no spatial or grid space.

1.3.3 Interaction

One of the main distinctive characteristics of agent-based models is that agents interact with other agents, the environment or themselves. Examples of agent-self interactions, where an agent interacts with itself, are reproducing a new agent, removing itself or modifying its own properties. Agent-agent interactions, where an agent interacts with another agent, include consuming an agent and possessing its resources, or sharing information with another agent. Agent-environment interactions occur when an agent alters its neighbourhood or when the environment has some effect on the agent. For example, when the agent observes the environment or moves within it (Wilensky & Rand, 2015). In addition, the environment also interacts with itself; "Environment-self interactions are when areas of the environment alter or change themselves." (Wilensky & Rand, 2015, p. 258). For example, the amount of resources might increase or decrease due to some natural fluctuations such as diffusion of a variable throughout the environment (Wilensky & Rand, 2015). These varieties of simple interactions produce a complex society (Wilensky & Rand, 2015). Notably, these interactions within ABMs constitute the basis for emergent

properties (Gilbert, 2007), which is what makes agent-based modelling such a powerful method.

1.3.4 Schedule

The schedule is the model operation order and time management of the commands. It starts with initialization procedure that creates the environment, agents and sets the parameters. The central part of schedule consists of a main loop that defines agents' actions, environment dynamics or other changes in one time unit of the model. There are different methods of updating the schedule in ABMs. The synchronous method is when all agents' states update at the same time. The method known as asynchronous is when states of some agents update before others (Wilensky & Rand, 2015).

1.3.5 Agents' Characteristics

Some of the key characteristics of agents include:

Heterogeneity

Each individual agent is explicitly operated by its preferences or its rules that may differ from one another.

Autonomy

Each agent is an autonomous entity, that observes its environment, decides independently to perform an action and interact with other agents that make changes in the environment.

Bounded Rationality

Bounded rationality concern agents' information and the computer power. Agents only have their neighbourhoods' information, not global information and they do not have endless computational power (Epstein, 2012). People should be modelled as bounded rational agents because of their limited cognitive abilities (Simon, 1957).

Reactivity

The ability of an agent to perceive the environment and perform action to the dynamics of the environment (Wooldridge, 2009).

Learning

Agent-based models are capable of modelling individual learning as well as population learning. There are different scopes of learning in agent-based modelling:

- Individual learning: Each agent can learn from itself through successful actions.
- Population learning: Agents with more effective actions are expected to reproduce more, and next generation will tend to improve.
- Social learning: Unlike individual learning and population learning, social learning builds on the agents' interactions. An agent is able to teach others (Gilbert, 2006).

1.3.6 Emergence

The behaviour of complex systems can be described in terms of individual agents' level, which is known as micro level, or in terms of the system as a whole, which is called macro level (Gilbert, 1996). Typically, an initial population of agents is released into the environment to observe a recognisable macroscopic social pattern. The interactions between individual agents in micro level generate a central social structures and group behaviours. Alternatively, the term emergent denotes to "stable macroscopic patterns arising from the local interaction of agents." (Epstein & Axtell, 1996, p. 35). In other words, emergence is a property that is generated at the macro level but is not specifically encoded at the micro level. Emergence is a characteristic of complex systems. It comprises of complex patterns that can often be generated from simple rules (Wilensky & Rand, 2015).

1.3.7 Reasoning Approaches in Artificial Intelligence (AI)

The nature of agents are associated with the reasoning methodology behind them. Agents as entities are able to perceive their environment via their sensors and act in the environment through their actuators (Russell & Norvig, 2014). A rational agent acts to maximize its performance based on its prior knowledge and the information it has perceived from the environment (Russell & Norvig, 2014). The intelligent agents' reasoning approaches are defined by their architecture in artificial intelligence. An agent architecture is a framework consisting of the elements from which agents are built and the methods of interaction with each other. The agent architecture determines the way in which the agent represents information, the way that the agent reasons, makes decisions and the action it takes to achieve its goal (Chin et al., 2014). Generally, there are three main categories of agent architectures in AI which include semantic architectures, cognitive architectures and classical architectures. The classical architecture is classified into four types including the logic-based architecture, reactive architecture, Belief-Desire-Intention (BDI) architecture and hybrid architecture (Chin et al., 2014). Figure 2 shows the hierarchy of the most dominant agent architectures in the literature.

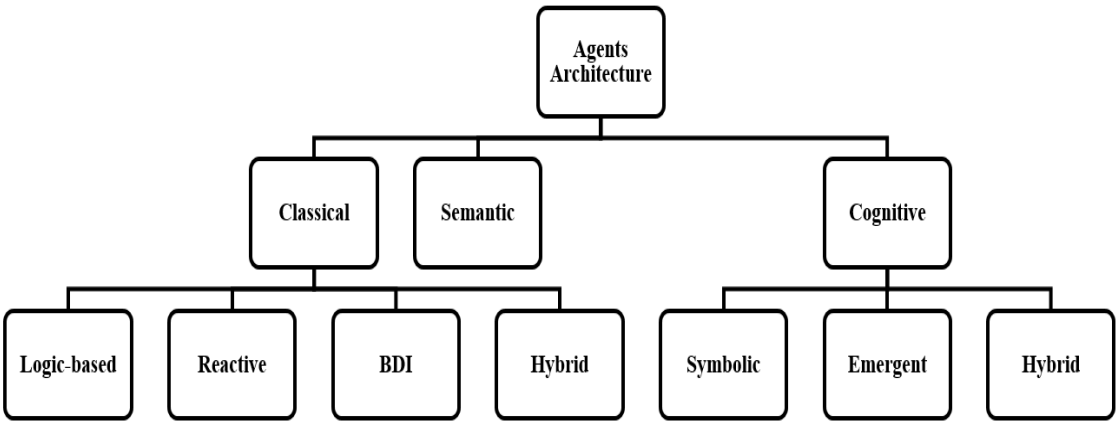


Figure 2. The hierarchy of agents' architecture

The semantic architecture involves semantic web technology. Using a logic-based (or deliberative) architecture is the classic approach for building intelligent agents in which the agent's action and reasoning is based on logical deduction and symbolic representation (Weiss, 2013). The hybrid classical architecture is the combination of reactive and deliberative agent architecture. There now follows a summary and discussion of reactive architecture, BDI architecture and cognitive architecture including symbolic, emergent and hybrid.

1.3.7.1 Reactive Architecture

The reactive architecture allows to construct agents that react to their environment with no reasoning skills. Agents' actions are based on mapping between stimulus and the response (Jones, 2008). Agents perceive the environment through their sensors and map the information they perceived to the one or more actions depending on the state of the environment. The best-known example of a reactive architecture is called the subsumption architecture which was developed by Rodney Brooks in the behaviour-based robotics research in 1986. In the subsumption architecture implementation the behaviour modules are considered as finite-state machines with no symbolic representation and symbolic reasoning (Brooks, 1986). Agents use a set of task accomplishing behaviour to make decisions. These behaviours map the perceptual information of the environment (situation) to actions. In addition, it is possible that many behaviours can be executed simultaneously. There is a mechanism that selects actions by applying a subsumption hierarchy for different behavioural modules into layers. The lower layers have a high priority and are able to inhibit higher levels which represent more abstract behaviours (Weiss, 2013). Inhibition is used to disable undesirable behaviours at a particular time or circumstance (Brooks, 1986). The subsumption architecture starts with a simple set of behaviours and then it is possible to extend it with the higher level of behaviours through further layers. This represents the

evolutionary design approach of the subsumption architecture (Jones, 2008). The subsumption layer is reactive and simple in nature. However, when the additional layers are added, they interfere with other layers and the control of the behaviours becomes an issue (Jones, 2008). The other development in reactive agents is Markov models; their behaviours are probabilistic in a dynamic environment. Markov models are widely used in stochastic processes modelling as well as AI environments in which a sequence of decisions must be made over time (Weiss, 2013).

Behaviour networks developed by Pattie Maes is another example of reactive architecture which is capable of selecting the best action for a given state of the environment. Behaviour networks consist of a network of behaviours with the related activation and inhibition links (Jones, 2008).

In general, one of the advantages of a reactive architecture is that they are simple, extremely fast and can be easily implemented in terms of hardware and software agents. Besides, through agents' interactions, it is possible that complex behaviours will emerge. In contrast, the disadvantages of reactive architecture include; firstly, the environment needs to be simple. Secondly, the lack of sufficient information regarding the agents' present states in their local neighbourhood to determine an acceptable action (Chin et al., 2014), makes the sequence of their actions problematic. Thirdly, the processing of agents' neighbourhood information restricts future planning and learning capabilities of agents (Chin et al., 2014) as reactive agents make decisions based on local information. Fourthly, it is difficult to construct agents which contain extensive number of layers. For example, it becomes difficult to understand the different behaviours and their relationships for more than 10 layers due to the dynamics and interactions between the different behaviours of the layers which become too complex to understand (Weiss, 2013).

1.3.7.2 BDI Architecture

The BDI architecture rests on the theory of human practical reasoning - the processes of deciding which action to take to achieve our goals - proposed by Michael Bratman in 1987. The concept of practical reasoning involves two key processes; firstly, "what goals we want to achieve" (Weiss, 2013, p. 28) which is called deliberation. Secondly, "how we are going to achieve these goals" (Weiss, 2013, p. 28) is known as means-ends reasoning. Beliefs, desires and intentions are the main mental state components of the BDI architecture. Beliefs represents the set of information an agent has regarding the state of the environment. These beliefs might not be necessary, correct or complete. Desires represents the agent's goals, motivations and what it wants to achieve. Intentions are a key factor in practical reasoning which lead to actions. Generally, an agent will not be able to reach all of its desires. Therefore, a subset of desires is selected to achieve them (filters). Intentions are these selective desires that agents have commited to achieve. One well-known example of BDI architecture which mimics the theory of human reasoning is called Practical Reasoning System (PRS). PRS data structure directly corresponds to beliefs, desires and intentions. Beliefs are the facts about the environments which are perceived via the sensors. A collection of pre-compiled plans (plan library) can be used by agents for the purpose of achieving different states of affairs depending on their desires and intentions. Desires represent a set of actions that agents should follow to achieve their goals. Each plan consists of a body and invocation condition (Weiss, 2013). The body of a plan is a set of actions that agents accomplish to achieve some particular state of affairs. The invocation condition defines the conditions whereby agents consider the plan. Thus, as the agent updates its beliefs through its interpreter, they continue to choose a plan which is consistent with the invocation condition and corresponds to the agents' active desires that act as their

intentions. Then, one action can be selected by the interpreter that represents the agent's present intentions and beliefs. Figure 3 shows a PRS representing of a BDI architecture.

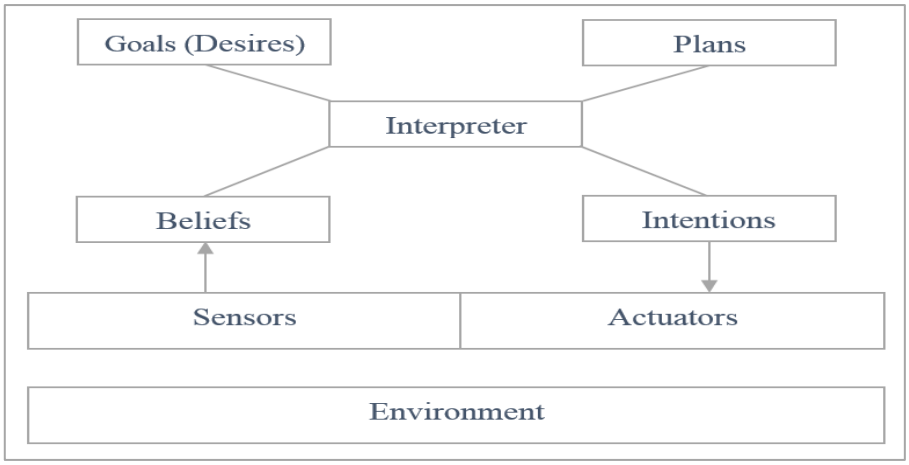


Figure 3. PRS in a BDI architecture for plan execution (Jones, 2008).

Although decision making processes become more stable through intentions, BDI requires to balance the intentions in different environments; in highly dynamic and unpredictable environments, BDI reconsiders its intentions often whereas in static environments less reconsideration is required (Weiss, 2013). One advantages of BDI architecture is that we intuitively understand beliefs, desires, intentions and the processes of choosing actions. In addition, the functional decomposition and subsystems for constructing an agent is clear. However, the problem is "how to efficiently implement these functions" (Weiss, 2013, p. 35).

1.3.7.3 Cognitive Architecture

The cognitive architecture is used to implement intelligent agents through emulating human behaviour and cognitive abilities. The essential structure and the processes of human minds and performances are specified through this architecture. The integration of artificial intelligence and cognitive science is crucial for exhibiting intelligent behaviour. For this purpose, a systems-level architecture is essential to support complex cognitive behaviours through a range of tasks. The underlying cognitive architecture is capable of storing perceptual knowledge and goals into memory, representing the stored information into mental structures and then operating functional processes on these mental structures,

including performance mechanisms and learning mechanisms (Langley, Laird, & Rogers, 2009).

Generally, based on two basic constituents of the cognitive architecture, memory and learning, three different subcategories are considered including symbolic, emergent and hybrid models. The symbolic architectures often apply a central control on information flow from sensors through memory to actions by using high-level symbols or declarative knowledge in a classical AI top-down, analytic approach (Duch, Oentaryo, & Pasquier, 2008). A popular example of cognitive symbolic architecture is SOAR (State, Operator And Result) where "a state is a representation of the current problem-solving situation; an operator transforms a state (makes changes to the representation); and a goal is a desired outcome of the problem-solving activity" (Laird & Bates Congdon, 2015, p. 5). Given each state can have only one operator at a time as SOAR runs, it uses the existing operator and selects the next operator until the goal has been achieved. The SOAR architecture comprises firstly a symbolic long-term memory that is encoded with the help of production rules for long-term knowledge to specify how to respond to different situations and secondly a short-term memory is structured as objects to store the information from sensors, current operators and current goals with properties and relations.

Emergent architectures, as a subcategory of cognitive architectures, specifies an explicit interaction between processing elements of network nodes in which their internal states change and a pattern emerges. Emergent architectures use a bottom-up strategy and are inspired by connectionist approaches. One example of this type of architecture is IBCA (Integrated Biologically-based Cognitive Architecture) which attempts to simulate brain functionalities by emulating the brain's high-level design. IBCA concentrates on posterior cortex (PC), frontal cortex (FC), and hippocampus (HC) regions of the brain which are responsible for the sensory and motor processing, dynamic and active memory and the fast

learning respectively (Duch, Oentaryo, & Pasquier, 2008). In order to provide a detailed structure of the modules, a large number of neurons is necessary to simulate cognitive functions. However, as the number of neurons increases, the issue of how well the emergent architecture simulates the cognitive function arises, this is referred to as the problem of scalability (Duch, Oentaryo, & Pasquier, 2008).

Hybrid architectures, another subcategory of cognitive architectures, are developed to integrate the strengths of two symbolic and emergent architectures into a more comprehensive cognitive framework. Symbolic architectures are reliable in high-level cognitive functions, such as planning and deliberative reasoning. However, problems arise in the formulation of symbolic entities from low-level information and managing large amount of information. In contrast, emergent architectures are capable of capturing the context-specificity of human performance and dealing with a large scale of low-level information. Yet, a weakness remains in capturing higher-order cognitive functions (Duch, Oentaryo, & Pasquier, 2008). Thus, a hybrid version can benefit from overcoming these limitations. Common examples of hybrid architectures are ACT-R (Adaptive Components of Thought-Rational), CLARION (Connectionist Learning with Adaptive Rule Induction On-line) and LIDA (The Learning Intelligent Distribution Agent).

The purpose of ACT-R is to simulate human cognitive tasks and to clarify the underlying mechanisms of perception, reasoning and action. ACT-R consists of a set of perceptual-motor modules, two memory modules, buffers and a pattern matcher. The perceptual-motor modules act as an interface between the system and the environment. The two memory modules include declarative memory (DM) and procedural memory (PM) for storing factual knowledge about the world and the way the system works respectively. The buffers act as temporary storage for communications between modules. The pattern matcher is for the purpose of finding a production in PM that matches the present state of the buffers.

ACT-R uses a top-down learning approach in a way that as a goal, factual knowledge or perceptual information appears, it becomes a chunk in the memory buffer. By applying Bayesian probability, the existing chunks that are used more frequently become more active and can be retrieved faster. Although, a large body of psychological research and intelligent tutoring systems have been implemented ACT-R architecture, still there is a lack of applications in reasoning and problem solving fields (Duch, Oentaryo, & Pasquier, 2008).

1.3.8 The use of ABM in social simulations

Initially computational modelling as a research technique started in the natural sciences from astronomy to biochemistry, it was neglected in the social sciences due to the lack of a computational approach that satisfied procedures and needs in the social sciences (Gilbert, 2007). However, from 1990, researchers started to realise that agent-based modelling features such as interaction, emergence, and micro and macro levels of phenomena are all applicable to the social sciences (Abrahamson & Wilensky, 2005). Since then, the number of studies in the social sciences applying agent-based modelling has increased dramatically. Agent-based modelling is a potential research tool for evaluating, formulating, training, predicting and understanding the procedures and the consequences of theories in the social sciences (Gilbert & Troitzsch, 2011). Agent-based modelling allows us to "grow" social structures in which certain sets of micro specifications lead to generating the macro phenomena of interest (Epstein & Axtell, 1996).

In addition, the recent surge in applying simulations, and in particular ABMs, in computational neuroscience and social cognition modelling has made remarkable advances in the field. Cognitive approaches and experimental psychological studies have progressed to provide a better understanding of social cognition rather the "pure" one by development of cognitive social sciences and exploiting ABMs (Sun, 2012). Macro-micro levels exist

in social-psychological interactions leading to complex phenomena in social and cognitive context (Tetlock & Goldgeier, 2000; Sun, 2006). For example, Marsella et al. (2004) implemented an agent-based model tool called PsychSim, for modelling interactions and influences between groups or individuals. Each agent is able to have its own decision making criteria based on its beliefs about the world and recursive models of other agents by applying theory of mind. PsychSim is capable of updating agents' beliefs based on other agents' actions and their psychological motivations (Marsella, Pynadath, & Read, 2004). Figure 4 shows a screenshot of PsychSim interface.

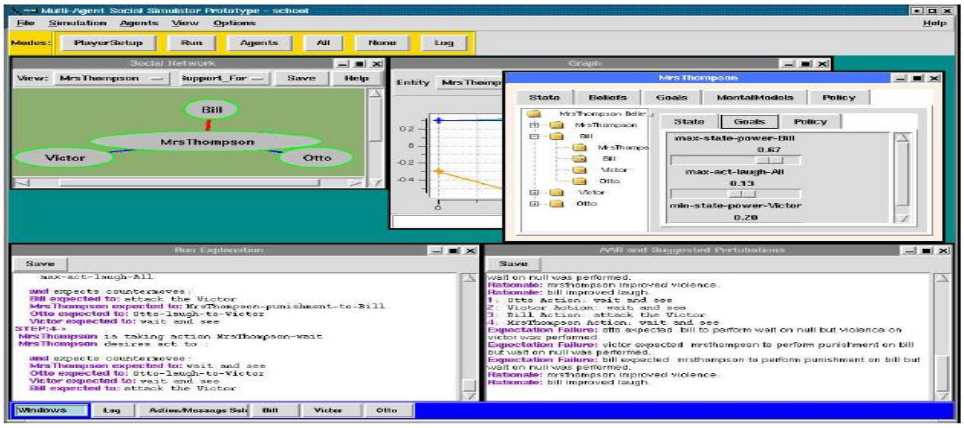


Figure 4. Screenshot of PsychSim interface

There are a variety of agent-based modelling applications in social simulations that relate to the topic of this thesis. One example is the Mod game, which was implemented by de Weerd et al. (2014). It simulates human behaviour as in rock-paper-scissors game with n-players. The Mod game has a mixed-strategy Nash equilibrium in which each action is played with equal probability. In a similar study, de Weerd (2013) and his colleagues have shown that the ability of using higher orders of theory of mind can be beneficial in some specific situations which will be discussed in section 3.5.

The opinion dynamics and bounded confidence model (Hegselmann & Krause, 2002), which demonstrates consensus forming within a group, is another example of considering psychological aspects in social simulations. In this model, decisions on the individual level

is based on simple behaviour rules, whereas the aggregated level results in complex and clustering patterns.

Another example is the attitude dynamics model (Brousmichea et al., 2016) which integrates a cognitive and an emotional component based on belief revision using a communication mechanism. The thesis now continues by explaining two of the most influential examples of ABMs in social simulations in more detail; segregation model and the game of life.

1.3.8.1 Segregation Model (Schelling's Model)

Thomas Schelling (1971) introduced a model for racial segregation in United States cities. Schelling proposed a threshold of tolerance to define when people are satisfied with the place they live in which a certain ratio of their neighbours belong to the same ethnic group as themselves. The model consists of blue and green agents representing two different racial household types which are placed randomly in a grid environment as an urban area. Thus, each cell on the grid can be either empty or occupied by a blue or green agent. Agents belong to one of the two groups and they are able to move to another cell if they are not satisfied with their neighbourhood. In other words, when the ratio of agents with the same ethnic group within eight cells in its vicinity (Moore neighbourhood), is less than the tolerance threshold then they are unsatisfied. Thus, the unsatisfied agents search for an empty cell in which they become satisfied and move there. The effect of agents' relocations, might unbalance the tolerance for the neighbours and cause some of them to become unsatisfied, resulting in a cascade of relocations (Gilbert, 2007). Thus, the initial random distribution of household segregation changes to patches of two groups, clustering together based on their group (Gilbert, 2007). These clusters indicate the racial prejudice could cause segregated patterns in cities (Schelling, 1971).

Figure 5 shows the segregation model with three different tolerance thresholds (F) in a city with a 0.2:0.8 ratio of group sizes. Three cases of integrity, mixed and segregation is demonstrated in this figure. (Hatnaa & Benensonb, 2012)

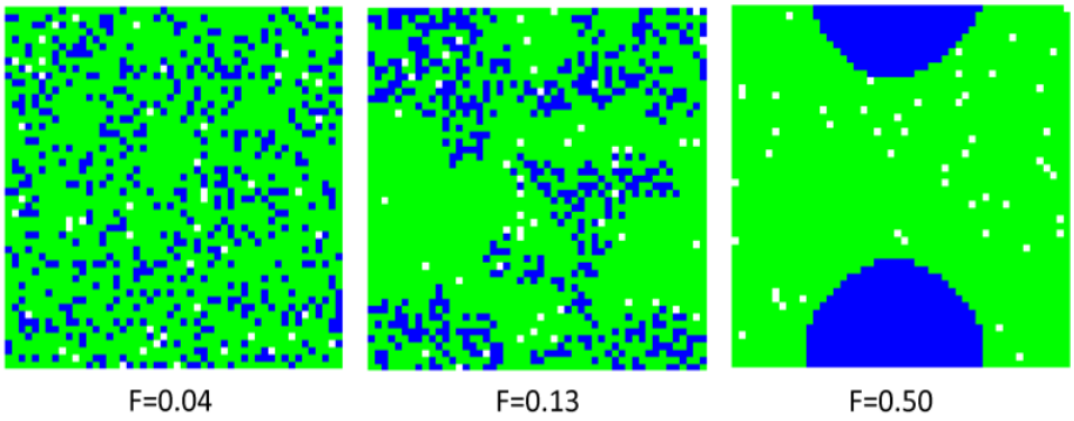


Figure 5. Integrated, mixed and segregated persistent patterns of the Schelling model. The patterns produced by the Schelling model for the 0.2:0.8 Blue-to-Green size ratio (Hatnaa & Benensonb, 2012).

The Schelling model is very simple to understand, simulate and analyse it. The emergence of clusters and segregation pattern is not predictable from the agents' micro level rules. In addition, the model can be tested with empirical data of cities and thus it has been an influential model in social simulations.

1.3.8.2 The Game of Life

The Game of Life, presented by mathematician John Conway in 1970, is a grid extending infinitely in all directions (cellular automaton). It is based on a few mathematical rules in which each cell can either be lit up and called alive, or, remain dark and is called dead. Whether a cell is alive or dead depends on its Moore neighbourhood, the eight cells in its vicinity. An alive cell dies of loneliness when there are either no neighbours or only one. Also, a cell with four or more neighbours dies because of overpopulation. A cell with two or three neighbours remain alive. A dead cell with three neighbours becomes a live cell. By using these simple rules, the game of life simulation generates various interesting

patterns depending on its initial conditions. The simple rules have a significant effect on the future patterns in the environment. In each time step, different shapes form and deform, a cluster of shapes move across the grid and reproduction happens with the new alive cells. Although the rules are simple the patterns which emerge are very complex. Figure 6 shows some of patterns of the game of life.

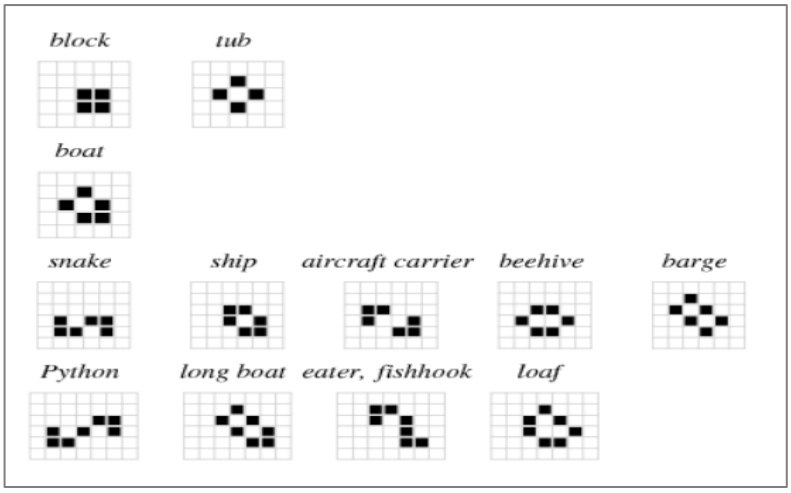


Figure 6. Patterns of the game of life

1.3.9 Simulation Software

The increasing demand of agent-based modelling for different fields has led to the development of more user-friendly agent-based modelling platforms such as NetLogo, AnyLogic, and Repast.

NetLogo (Wilensky, 1999) is a general purpose agent-based modelling language which uses the Logo language syntax as a basis and includes a library of models in different fields. NetLogo is a high-level platform, with built-in graphical interface capable of reducing the programming effort, providing a simple but powerful programming language (Railsback, Lytinen, & Jackson, 2006). However, it lacks the flexibility of a standard programming language due to limitation of the control and structuring capabilities.

The AnyLogic platform enables modelling complex phenomena by using a graphical user interface and Java code. The object-oriented paradigm which is supported by AnyLogic makes it a potential option for this study. However, the cost of the AnyLogic Research License is high. For the purpose of this project, Repast Simphony has been selected; a popular open source Java-based modelling platform that is capable of modelling complex systems such as theory of mind. Figure 7 shows the interface of Repast Simphony.

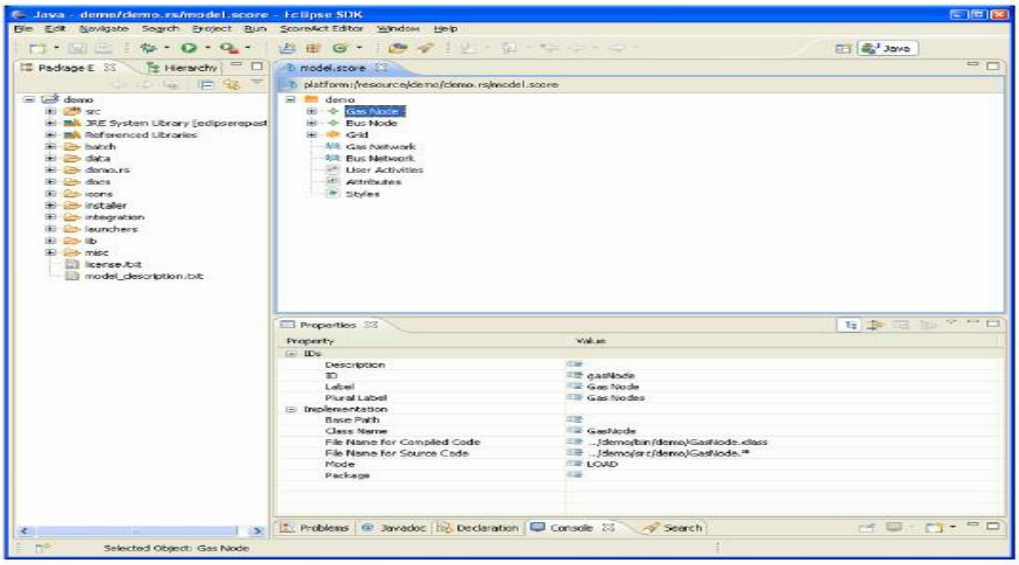


Figure 7. The simulation environment in Repast Simphony

1.4 Why does this thesis apply ABM for theory of mind?

Agent-based modelling was selected as a methodology for the purpose of theory of mind simulation for several reasons. Firstly, the complex nature of theory of mind requires a powerful empirical research technique that naturally associates with the complex systems. One prominent technique for complex systems is ABM that enables the interaction between the parts of the system. This viable characteristic of ABM utilises the agents' interactions to pursue their goals in the environment. More importantly, it facilitates the agents' mental states anticipation or manipulation behind this interactions in the simulation of theory of mind in this thesis. Thus, ABM is a coherent research laboratory for the concept of theory of mind.

Secondly, despite the large amount of important research in theory of mind literature, there is still ambiguity in research methodology and a lack of standard experimental methods in the field. ABMs are able to provide a promising approach to reconstruct and simulate our knowledge of theory of mind to understand its patterns and characteristics in the society. The agent-based models in this thesis attempt to imitate the concepts of understanding others' belief and desires, based on the developmental literature, and explore the main patterns, efficiency and the process of the concept.

Thirdly, the analysis of theory of mind procedure into a systematic approach is not directly feasible and there is no natural way for studying it as a social behaviour. However, ABMs as interdisciplinary approach are able to explore theory of mind concept from individual and social aspects simultaneously. Besides, ABMs are capable to serve as a proof of the underlying mechanisms of a concept (Sun, 2006) such as theory of mind.

Fourthly, ABMs are capable of constructing and reconstructing concepts and procedures such as reasoning through simulation in a social context. Thus far, this has not proved possible through practical experiments with people. Therefore, agent-based modelling offers a reliable alternative for researchers in theory of mind field.

Finally, developing an agent-based model to explain theory of mind features, from micro to macro level, enables us to manipulate the parameters and understand the effect of the changes on behaviour of the system. Moreover, ABMs act as a simplified model of reality, parallel to other modelling tools, capable of predicting the behaviour of the system and providing analogous algorithms.

In general, it is not possible to study the mental processes merely by behavioural experiments; the complexity of human mind together with its manifestation in the flexible behaviour requires computational models to analyse the complex detail (Sun, 2008).

ABMs as computational models are not restricted by mathematical equations, thus they simultaneously are expressive and precise (Sun, Coward, & Zenzen, 2005). Agent-based models not only hold flexibility and process-based characteristics, they also express both individual level and social aspects of the phenomenon. Besides, they are capable of showing the dynamics, interactions and parts in theory of mind processes. Based on these reasons, agent-based models are the most advantageous fit for theory of mind simulations.

CHAPTER 2

2. AN AGENT-BASED MODEL FOR BELIEF REPRESENTATION

2.1 INTRODUCTION

Two children are playing hide and seek in the jungle. The hider child first makes sure that the seeker does not see her as she is hiding. Why does she do this? How does she do this? This example reveals that children are able to track others' field of view. There is a realistic link between children's abilities to play games, such as hide and seek, successfully and applying false belief reasoning (Wellman, 2014).

In everyday life, a large amount of belief processing, true and false beliefs, occurs in our mind. Yet, in order to efficiently build connections and communicate with others even in a small and simple group of people, it is necessary to distinguish others' true and false beliefs. The inability to infer and understand others' beliefs has been identified as one possible reason for deficits in social life interactions in individuals with Autism Spectrum Disorder (ASD) (e.g. Frith, 2001; Roekel, Scholte, & Didden, 2010).

One primary way to analyse belief representation is to distinguish between others' true and false beliefs. The best presumption about others' beliefs in everyday experiences is that it is the same as one's own because peoples' ordinary beliefs are usually true and true beliefs are default beliefs (Leslie, Friedman, & German, 2004). Thus, understanding others' true beliefs is a relatively straightforward process that involves a shared belief between one and others.

In contrast to true belief, understanding others' false belief is more demanding on cognitive resources such as memory and reasoning (Apperly et al., 2007). In addition, as already explained in general introduction section 2.1 , Leslie et al. (2004) suggest that to succeed in false belief task, it is necessary to inhibit the true belief default such that an alternative belief with different content can be selected. The objective of false belief tasks was to examine children's ability of inferring others' perspective, which was different from the real world state. This might be the underlying reason that understanding others' false belief

is considered as an acid test for presence of theory of mind ability (e.g. Wellman & Bartsch, 1988; Workman & Reader, 2014; Doherty, 2009).

In contrast, Bloom and German (2000) are the principal opponents for considering false belief tasks as an acid test for theory of mind. Their critical paper explains two reasons that false belief tasks needs to be abandoned as a test for theory of mind. The first reason given is that to pass a false belief task it requires abilities other than theory of mind. Secondly, theory of mind does not require the ability to reason about false beliefs. This will be discussed in section 4.9.

In a similar vein, minimal theory of mind proposed by Butterfill and Apperly "enables those with limited cognitive resources" " to track others' perceptions, knowledge states and beliefs" without "representing propositional attitudes, or any other kind of representation, as such." (Butterfill & Apperly, 2013, p. 1). They suggest four principles for minimal theory of mind to be able to track others' perceptions and belief-like, which has been explained in the general introduction section 1.2.6. Apperly and Butterfill claim that one with minimal theory of mind ability is able to pass the tests which are supposed to be an acid test of theory of mind, as well as many false belief tasks.

In addition to the lack of consensus on an acid test for theory of mind, the diversity of false belief task design has been rapidly increasing. A recent review by Schaafsma et al. (2015) highlights that there are more than 36 different variety tasks such as false belief versus false photograph, story-based format for false belief, false belief versus true belief, false belief and subjective preference and false belief versus physical reality is employed in fMRI studies (Schaafsma et al., 2015). The design of false belief tasks sometimes contains ambiguity or complexity, which makes it difficult to accurately interpret the experiments results. Historically, the literature has expanded with the diversity of tasks and with very little consensus on core principles. This might be a practical motivation to address the

question that which sets of basic processes are shared across the different varieties of false belief tasks.

This chapter presents a computational model of belief representation to address some of these inconsistencies in the literature. Besides, it explores some of the advantages and costs of understanding others' beliefs. For this purpose, an agent-based model called "Belief Representation Model" (BRM) was designed to shed light on false belief's processes at both the micro and macro levels. On the micro level, BRM examines belief representation concept, procedures and the minimum resources it might require in a dynamic environment. On the macro level, the aggregated results of BRM are compatible with the empirical effects of passing or failing false belief tasks in a virtual society; the BRM simulation results reflect the effect of understanding others' belief in agents' performances.

The Martin and Santos (2014) experiment, which has been outlined in general introduction section 1.2.8, formed the underpinning premise of BRM. The analysis of BRM will lead to a systematic approach to understand others' beliefs, which determines the necessary shared process for belief representation. Building upon on their experiment, BRM has involved the design of two types of simple agents with different abilities of belief representation. One type of agent in BRM follows the standard false belief task procedure and understands others' beliefs whereas the second type of agent is not capable of understanding others' beliefs. This will be explained in detail in section 2.4.3. Furthermore, on the macro level, different agents' performances will be assessed and analysed. This performance assessment highlights the efficiency of different theory of mind abilities.

Moreover, the link between minimal theory of mind principles and the proposed systematic approach for belief representation will be described.

2.2 BRM METHODOLOGY

An agent-based model is designed to simulate the process of others' true and false beliefs and to explore the successful procedure of belief representation. The model consists of 3 types of agents, representing different capabilities of tracking others' beliefs in a virtual world.

2.2.1 Environment

The virtual environment is made of a grid space of 50 by 50 in which agents interact within the environment based on their rules to achieve food. The space is toroidal meaning that if agents move to one border of the grid; it appears on the opposite border.

Time Step

The time measurement is called a tick; it is a step in the simulation when agents simultaneously perform their actions depending on their rules. The default number of time steps for the simulation is 1000 ticks.

Neighbourhood

The agents' first neighbourhood refers to the Moore neighbourhood which consists of eight cells around the agent's cell that touch it (Wilensky & Rand, 2015). The second and third neighbourhoods consist of 24 and 48 cells respectively. Figure 8 illustrates different levels of neighbourhood for the red agent located in the green cell; the first neighbourhood consists of yellow cells, the second neighbourhood consists of yellow and grey cells whereas the third neighbourhood includes yellow, grey and blue cells. The concept of agents' neighbourhood defines two important features in the simulation: field of view and field of movement.

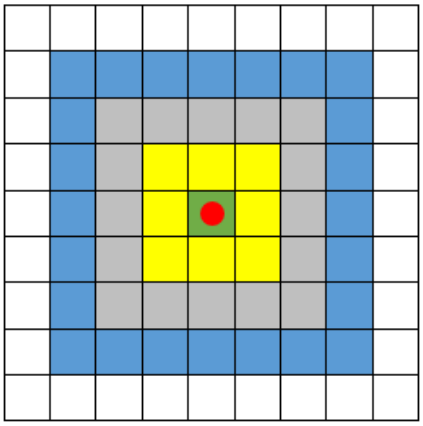


Figure 8. Agents' different neighbourhoods
The first neighbourhood (yellow cells), the second neighbourhood (yellow and grey cells) and the third neighbourhood (yellow, grey and blue cells) of the red agent in the green cell.

Field of view

The agent's field of view is the observed environment which consists of the cells within its neighbourhood.

Field of movement

An agent's field of movement includes the eight cells around the agent and the cell it currently occupies, as is shown in Figure 9. These are the cells to which the agent may move.

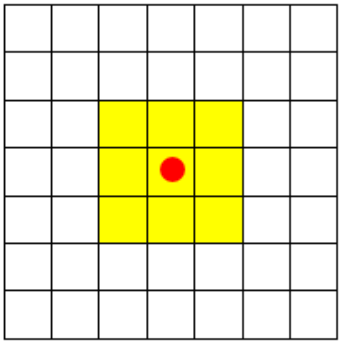


Figure 9. The area of field of movement
The yellow area shows the field of movement of the red agent

Food

Food is indicated by green cells and agents are required to consume food. Agents' ultimate goal is to consume as much food as possible. Therefore, the number of consumed food by

agents is a criterion for the performance measurements. The number of food parameter defines the number of available food in the environment in each time step.

At the start of each time step, the food that was not used by any agent remains in the environment. While the food, which was used by agents, will be replaced randomly in the grid for the next time step. Therefore, the number of food in the environment remains constant through the simulation time steps.

2.2.2 BRM General Rules

The model consists of 3 types of agents interacting within the environment: Monkey, Infant and Control agents. For ease of reference, Monkey, Infant and Control agents are initialised in capital letters. Note that the Monkey and Infant agents' names are not based on similarity with monkeys and infants in the real world. However, the main motivations for these names are based on Martin and Santos (2014) experiment and the idea that there are two types of agents with and without belief representation capability. In other words, Infant agents represent the ability to understand others' false beliefs whereas Monkey agents are unable to have others' belief understanding competence.

Each agent has visual perception from its own field of view. Agents encounter food in their field of view and correctly register their locations. Each time step agents move in a random turn. BRM class diagram, is shown in Figure 19, this represents the agents, their main functionality and the links with the environment.

- At the start of each simulation run, the initialisation procedure will be run. A defined number of agents (for each type) and a defined number of food are randomly placed in the virtual world.
- Agents move to an empty cell within their field of movement.
- If there is food in the field of movement, agents move towards the food.
- Agents can move to a cell that is not occupied by other agents.

- If there is no food in the first neighbourhood, agents move to a random cell depending on the agents' strategies.

Parameters

The BRM simulation analysis is based on modifying two parameters; number of agents and number of food.

Actions

Agents move to the cells that contain food as their targets. However, their movement depends on their abilities, rules and strategies on the micro level.

2.2.3 Agents' Strategies

Agents possess different abilities in terms of collecting information from their field of view and understanding others' actions as they do not collect all of the information in their field of view. Then, agents process the information they have collected. Thus, each type of agent acts based on its particular ability and strategy as follow:

2.2.3.1 Monkey Agents' Strategy

The general diagram of Monkey agents' strategy is illustrated in three processing directions of collecting information, recording information and action in Figure 10. Monkey agents are able to see the first neighbourhood as their field of view and collect the information about the location of the food they can consume, which is shown as collecting information in Figure 10. They do not reason about which food to choose, rather they randomly select the food.

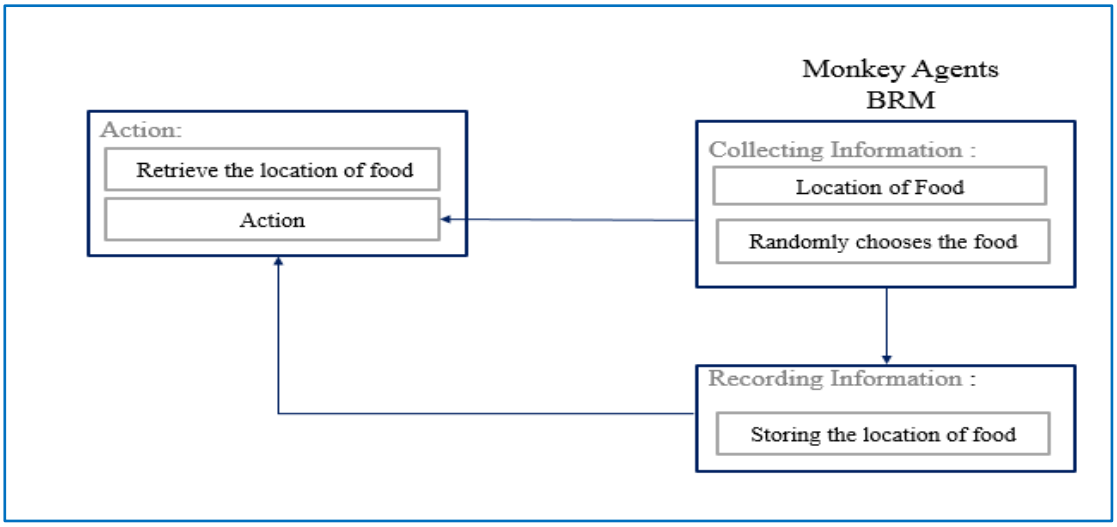


Figure 10. Monkey agents' arrow and box diagram

Monkey agents can remember the location of the food from the past time step by storing it in their memory, that is demonstrated as recoding information in Figure 10. Therefore, when food is available their action is consuming the food whereas, once there is no food available, they are able to retrieve the location of the food from the previous time step and move towards it, which is shown as action in Figure 10. In addition, Figure 11 depicts their strategy of retrieving the location of food in more detail. At the time step t, as Figure 11.a shows the yellow Monkey agent encounters and stores green and black food (In this situation, Monkey agent always moves randomly towards one food). It randomly chooses to consume the black food but stores the location of the green food in its memory. Thus, at the start of time step t+1, the yellow Monkey agent, which has stored the location of the green food, has no access to any food in its field of view as is demonstrated in Figure 11.b. Based on its rules, it will come back towards the green food. Figure 11.c depicts that at the end of time step t+1, the yellow Monkey agent moves to its previous location, towards the green food (as the location of the green food is stored in its memory from the last time step).

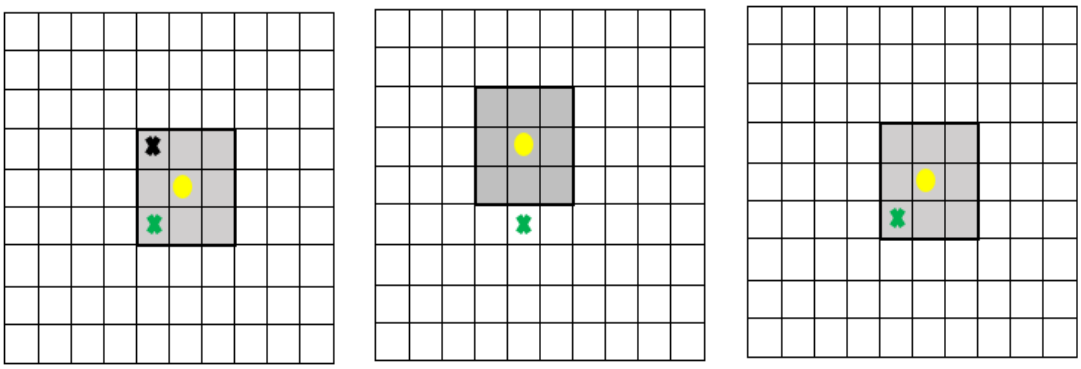


Figure 11. The steps in which Monkey agents retrieve the previous food information.
a) Monkey agent (yellow circle) moves towards the black food and stores the location of the green food in its memory at time step=t.
b) Monkey agent has no access to food in its field of view (grey cells). It remembers the location of the green food at the start of time step t+1.
c) Monkey agent moves to its previous location towards the green food in time step t+1.

Monkey agents are able to recognise their own false beliefs over time; when they move towards a restored food and it is no longer is there. Nevertheless, they do not collect any information about other agents' field of view and they are unable to take others' perspective into account.

2.2.3.2 Infant Agents' Strategy

Infant agents are able to collect information about the location of food and other agents' perspective regarding the food. They are able to reason which information to collect and use. Infant agents are able to consider Monkey agents' beliefs regarding the location of the food. In order to do so, firstly, Infant agents search in their first neighbourhood and collect information on the location of the food. Then, the Infant agents identify all Monkey agents in their third neighbourhood. Infant agents are able to see the first, second and third neighbourhood as their field of view. Moreover, Infant agents identify the food that is in each Monkey agent's field of view which also exists in its own third neighbourhood. Note that the Infant agents only use the third neighbourhood to identify the Monkey agents' perspective. This extension allows Infant agents to consider a reasonable numbers of

Monkey agents' perspectives. Infant agents choose the Monkey agents and the information from their perspectives when they both have access to the same food at the current time step. Infant agents make decisions to move towards a food based on two types of information:

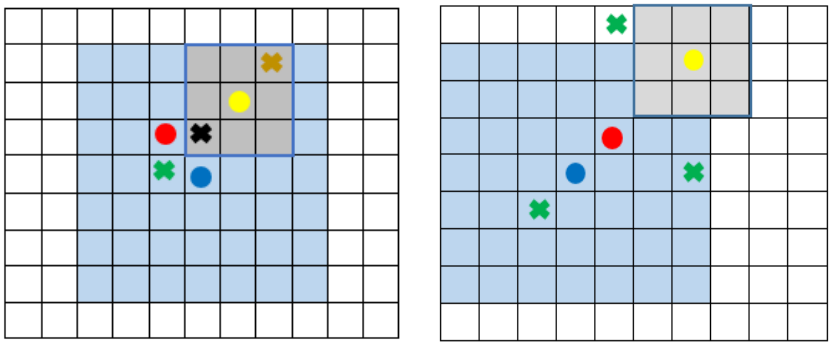
a) The previous time step information about Monkey agents' perspective regarding the location of the food.

b) The current time step information about the real world.

As Infant agents track others' field of view and store Monkey agents' perspective in their memory, their first priority is creating false beliefs for Monkey agents by retrieving the information about Monkey agents' perspectives. There are two cases where the Infant agent recognises Monkey agents' false beliefs:

Case I: Another agent consumes the registered food by the Monkey agent.

Suppose Figure 12.a is the start of time step t, and the Infant agent (blue circle) encounters two food (green and black), and simultaneously stores the Monkey agent's (yellow circle) perspective of the location of black and brown food.



(a) (b)

Figure 12. Infant agents' strategy regarding Monkey agents' false beliefs

a) The Infant agent (blue circle) encounters green and black food whereas Monkey agent (yellow circle) encounters the black and brown food in their field of view in time step t. The red agent is another agent that has access to the black and green food.

b) Red agent, Infant agent and Monkey agent respectively consume the black, green and brown food. In Monkey agent's perspective, the black food is still in its previous location. Infant agent memorises this perspective. Thus, the Infant agent identifies Monkey agents' false beliefs as another agent consumes the food registered by the Monkey agent.

Figure 12.b illustrates time step t +1, in which the red agent (can be Infant or Monkey agent) consumed the black food. Whereas, the Monkey agent and the Infant agent consumed the brown and green food respectively. The Infant agent predicts that the Monkey agent will return to its previous position for the black food in the next time step (t+2), and encounters a false belief situation, noticing that the Monkey has no alternative food. At this stage, the Infant agent identifies the false belief of the Monkey agent.

Case II: Infant agent consumes the registered food by Monkey agent.

One of the Infant agents' main rules is to prioritise consuming the food that the Monkey has registered as an alternative food in its memory. Thus, as Figure 13 demonstrates, the Infant agent (blue circle) moves towards the black food to create a false belief situation for the Monkey agent (yellow circle).

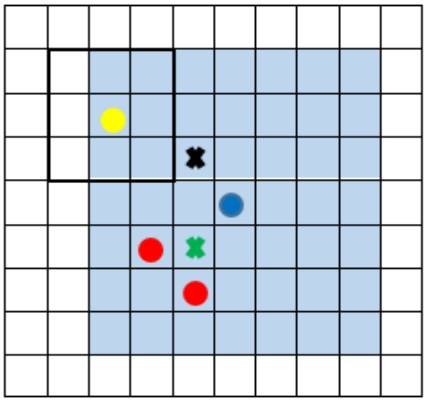


Figure 13. Infant agents' strategy

The Infant agent (blue circle) encounters two green and black food. Note that the Infant agent has already stored the Monkey agent's perspective regarding the black food location, thus, the Infant agent's priority is to move towards the black food to create a false belief for the Monkey agent (yellow circle).

As already stated, unlike Monkey agents, Infant agents are capable of understanding the Monkey agents' false beliefs. In order to do this, Infant agents need to inhibit their own perspective, which they normally use. The Infant agents' perspective which includes the real location of the food has to be replaced with the Monkey agents' perspective. The

location of the food in the Monkey agent's perspective is different from the Infant agent's own perspective when the Infant agent identifies the Monkey agent's false belief.

In a situation where there is more than one food, Infant agents plan and choose the food that creates a false belief for Monkey agents. Infant agents prioritise the food based on two conditions. Firstly, that the food has previously been stored in the Monkey agent's memory and which in the Monkey agent's perspective is still there. Secondly, that the Monkey agent has no alternative food in its field of view. In other words, Infant agents choose the food that could create false belief for Monkey agents, this is called priority function. In cases where there is no Monkey agent interested in the food, the Infant agent will choose the food which has the maximum number of agents around it. In fact, it uses the most vulnerable food because this might result in more competitors having no access to the food. Thus, for each food in their field of view, Infant agents calculate the total number of competitors at the current time step. Figure 14 illustrates an activity diagram of Infant agents including priority function. In addition, Figure 15 illustrates the general diagram of Infant agents' strategy based on collecting information, recording information, retrieving information, reasoning process and expressing others' belief-desire phases, which will be elaborated later in the discussion.

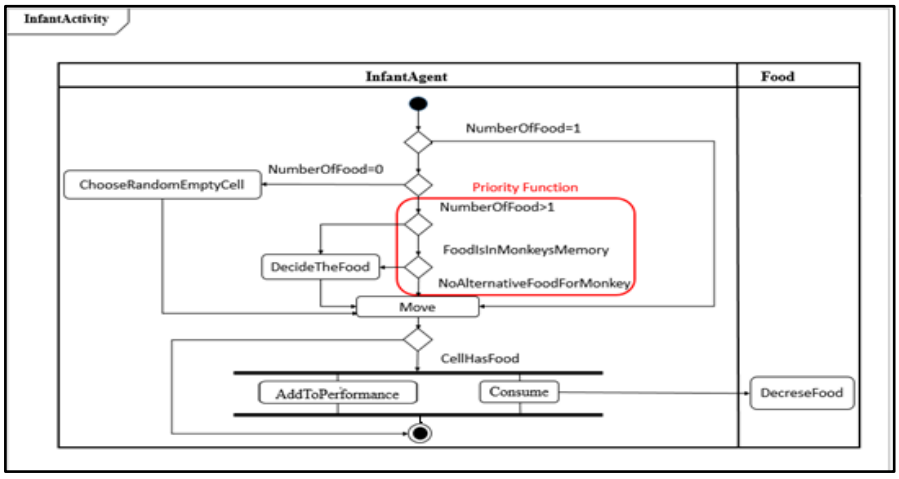


Figure 14. An activity diagram of Infant agents representing Priority function

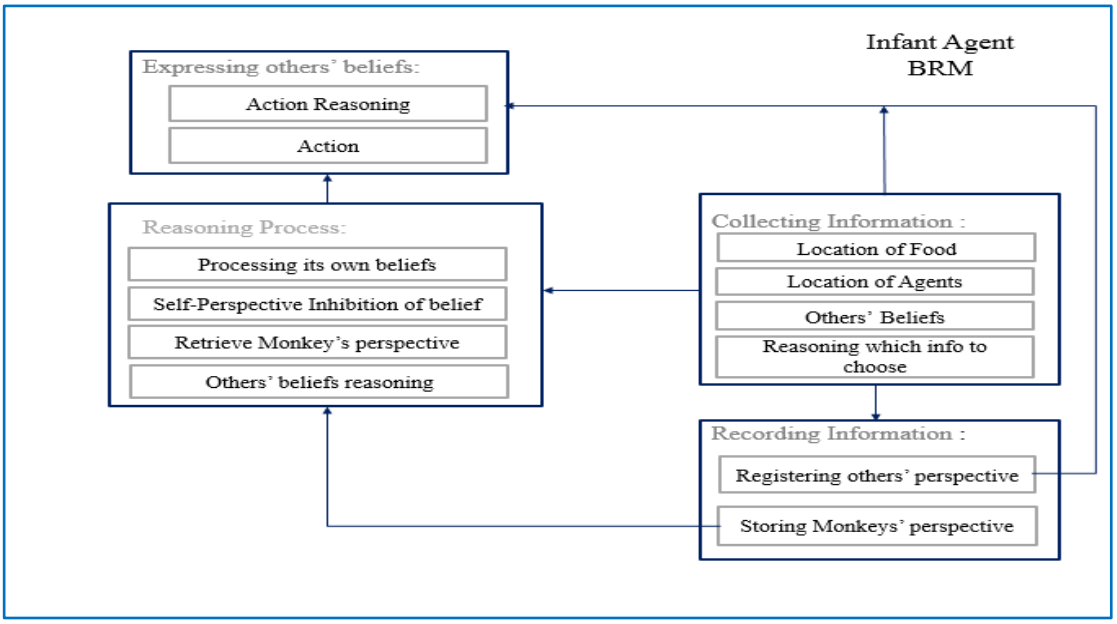


Figure 15. Infant agents' arrow and box diagram

2.2.3.3 Controls Agents' Strategy

Control agents have been introduced as a control measurement. They collect information from their field of view regarding the location of food and other agents' perspectives. Control agents are able to track others' field of view. Control agents partly act similar to the Infant agents but they do not store Monkeys' perspective information for using in the next time steps. In situations where there is more than one food available, Control agents calculate the total number of competitors which have access to the food and move towards the food with the higher risk of being consumed. For example, the Control agent (red circle) in Figure 16 encounters two green and black food. The black food has one competitor (yellow agent) whereas the green food has two other competitors. Thus, it selects to move towards the green food. Control agents do not store Monkey agents' beliefs information. They use the present time information about the world (part b of Infant agents' information which has already explained in Infant agents' strategy section) and have no access to the Monkeys' past perspective information (part a of Infant agents' information). Therefore,

Control agents' performance is a reliable measurement because they are able to track others' field of view at the current time, which is partly similar to the Infant agents, and also their field of view is wider than the Monkey agents.

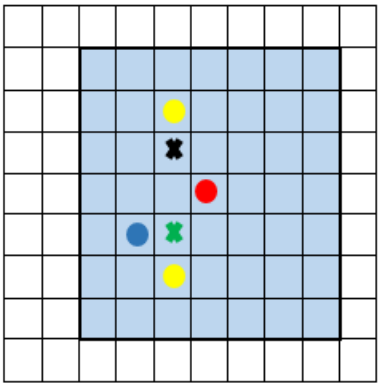


Figure 16. Control agents' strategy regarding competitors

Control agent (red circle) encounters two green and black food. It selects the food with the maximum number of competitors. It moves towards the green food with two competitors rather than the black food with only one competitor.

Figure 17 shows the general diagram of their strategy based on collecting information, registering information and action procedures.

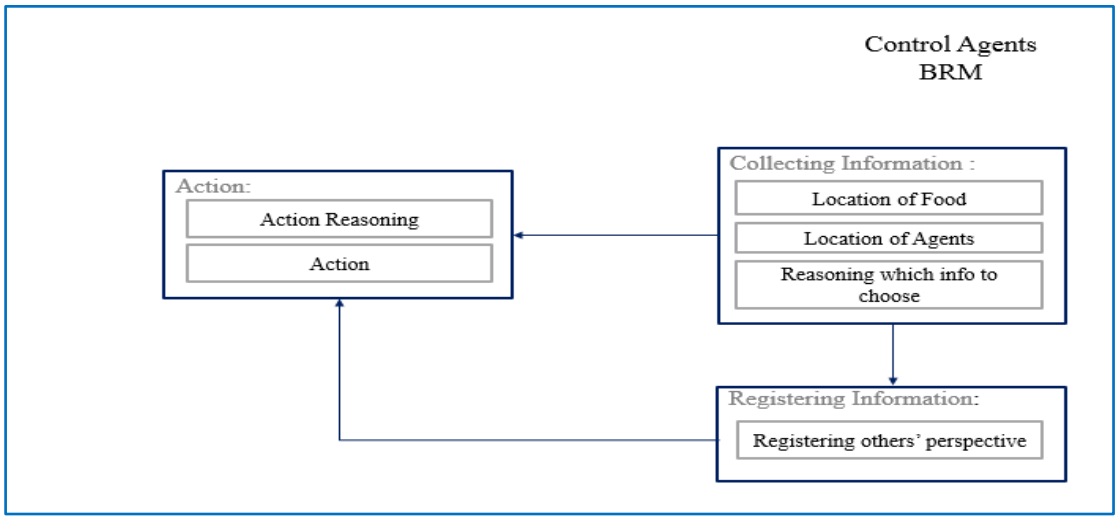


Figure 17. Control agents' arrow and box diagram

2.2.3.4 A comparison between agents' abilities

There are certain preconditions for agents which enables them to understand others' beliefs in BRM. Firstly, it is essential that agents are capable of tracking which one of the food

can be seen by other agents. Agents need to store the information about which food has been seen by which agents. In other words, sub-abilities such as tracking others' field of view and registering others' perspective of the location of the food are the starting points of understanding others' beliefs. It is essential to retrieve the information regarding which agents had seen which food in the previous time steps.

Table 1 shows Monkey, Infant and Control agents' abilities; Infant agents possess all three abilities of tracking others' field of view and registering their own and others' perspective. Thus, Infant agents meet the preconditions of passing the false belief task. In contrast, Monkey agents only store their own perspective of the location of the food whereas Control agents are only able to track other's field of view at the current time step.

Agent	Track others' field	Store their own information	Store other's perspective	Return to the previous food	Registering others' perspective
Monkey		✓		✓	
Infant	✓	✓	✓		✓
Control	✓				✓

Table 1. Agents' different sub-abilities

All agents are able to use their vision ability; they are capable of seeing the local environment. Moreover, Infants and Control agents are able to track others' field of view, while Monkey agents remember the food from the last time step. Infant agents are capable of recording and retrieving the information from Monkey agents' perspective. They are able to inhibit their own belief about the location of food and use the Monkey agent's perspective. In the case where this information is not similar to the current reality of the world, they can identify Monkey agents' false beliefs. Although Control agents are able to track others' field of view but they are unable to store others' perspective regarding the location of the food. Control agents are not designed to inhibit their own perspective; therefore, they are not capable of passing the false belief task.

2.2.4 How do Infant Agents understand Monkey Agents' false beliefs?

Firstly, Infant agents collect information about the location of the food in their field of view. Then, Infant agents track Monkey agents' perspective about the location of the food. Secondly, Infant agents require memory to store the information about Monkey agents' perspective about the registered food (in time step t-1) where the shared area of their field of view contains the food. Thirdly, Infant agents collect the current time step information about the location of the food (time step t) which no longer exists in Monkey agent's field of view.

Infant agents inhibit their own perspective and thus, do not use the real information about the location of the food, which is the same as their own perspective. Infant agents retrieve the stored information about Monkey agents' perspectives, which are in their field of views. Infant agent temporarily ignores its own beliefs in regards to the location of the food and uses the Monkey agents' beliefs. This is considered as self-perspective inhibition of Infant agents. Thus, the information processing continues based on retrieval of the Monkey agents' perspectives information rather than the real information on the location of the food and the Infant agents' perspective.

Given the current information on Monkey agent's location (time step t), and Monkey agent's perspective regarding the food (from time step t-1), Infant agents reason about Monkey agent's desire towards the food. Unless there is an alternative and closer food available for the Monkey agent, Infant agent considers that Monkey agent's intention is to move towards the food.

Although, when the registered food by the Monkey agent is consumed by another agent, the Monkey agent still moves towards the food; this is because the Monkey agent believes that the food is still in that location. This condition is counted as Monkey agents' false belief. This process enables Infant agents to recognise Monkey agents' false beliefs. Infant

agents are able to detect Monkey agents' false beliefs as they store Monkey agents' perspective simultaneously.

The false belief representation in BRM, hinges on three different perspectives; Infant agent perspective, Monkey agent's belief, Infant agent perspective of Monkey agent's belief. In time step t, all of these perspectives are identical. However, in time step (t+1), there is a contradiction between Infant agent's perspective about the location of the food and Infant agent's perspective of Monkey agent's belief. Thus, one critical parameter is the concept of time steps in BRM. This description is concisely depicted in Figure 18.

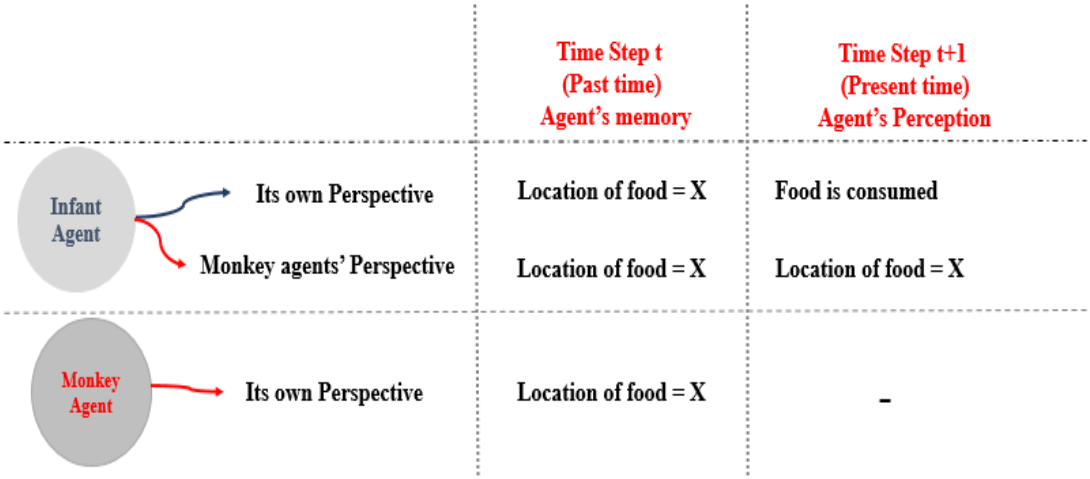


Figure 18. Agents' perspectives in false belief situations

False belief concept based on Infant and Monkey agents' perspectives in BRM, where X is the location of the food.

2.2.5 BRM Implementation

UML (Unified Modelling Language) is used to provide a visual version of the BRM design. UML is a standardised object oriented graphical modelling language with hundreds of modelling symbols for composing different kinds of diagrams (Unified Modelling Language, 2015). Class, sequence, state and activity diagrams are the most useful ones for ABM development (Bersini, 2012).

The BRM design consists of a society with two different type of agents moving to consume food in a grid space, in which agents and food have been placed. Figure 19 illustrates the BRM class diagram including the classes (which are basically templates for objects) and the relationships between them. The Society class is a composition of the Agent, Food, and Grid class. There is a composition association (the black diamond sign) between Society class and other classes indicating that Society class is a container for objects of other classes. Thus, when the Society ceases to exist at the end of the simulation, all of the objects inside it will be destroyed.

The Agent class is an abstract class, inheriting from two sub classes, Monkey Agent and Infant Agent. The two subclasses of Agent class including Monkey Agent class and Infant Agent class, inherit Agent class methods. However, they have their own specific methods and are also able to redefine the methods of their superclass. For example, Infant Agent class has three additional methods including Prioritise, Reason and Record in addition to Move and Consume methods which are inherited from its superclass. Although, the Record method exists in both of the subclasses, they are different methods and have different parameters. The Record method in Infant Agent class, stores the information about Monkey agents as well as the food whereas this method in Monkey Agent class only has one parameter regarding the food. There is a 1-1 association between Agent and Grid indicating that each agent is placed in one and only one cell of the grid. Similarly, there is a 1-1 association between Food and Grid demonstrating that each food can exist in one cell.

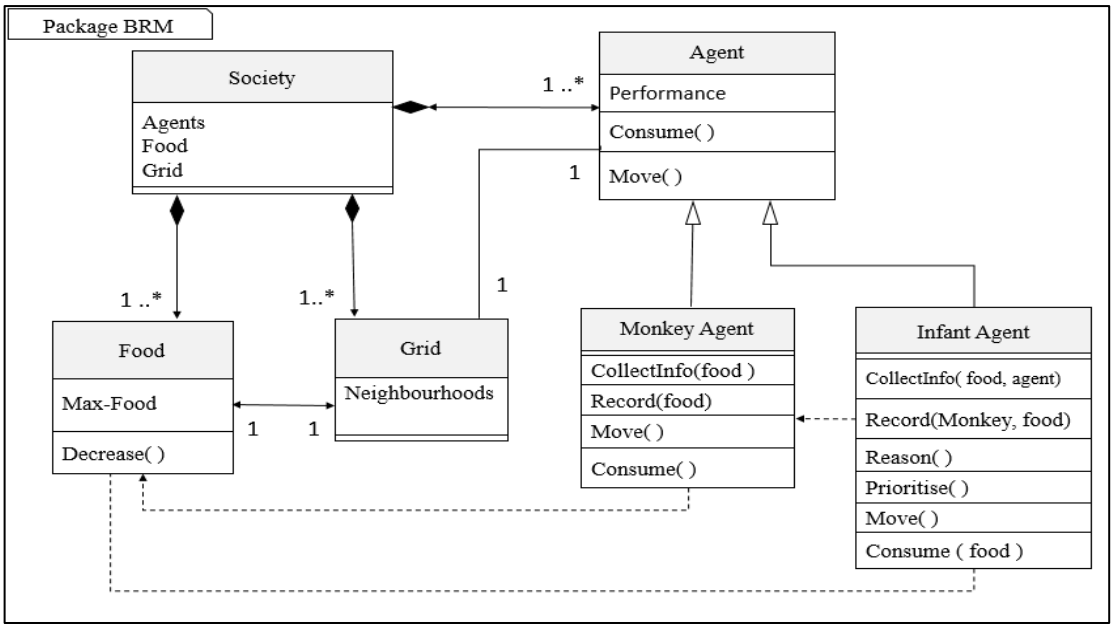


Figure 19. Class Diagram of BRM

There is a dependency relationship from the Monkey Agent class to the Food class. This relationship shows that the Monkey agent consumes the food by passing the information about the food, as a parameter, to Consume method. A similar relationship exists between the Infant Agent class and the Food class. Also, there is a dependency relationship from the Infant Agent class to the Monkey Agent class. The Move and Consume methods are the same for both of subclasses. The Food class includes the Decrease method that as the food has been consumed by an agent, it will be removed from the grid. Figure 20 shows a screenshot of running BRM in action.

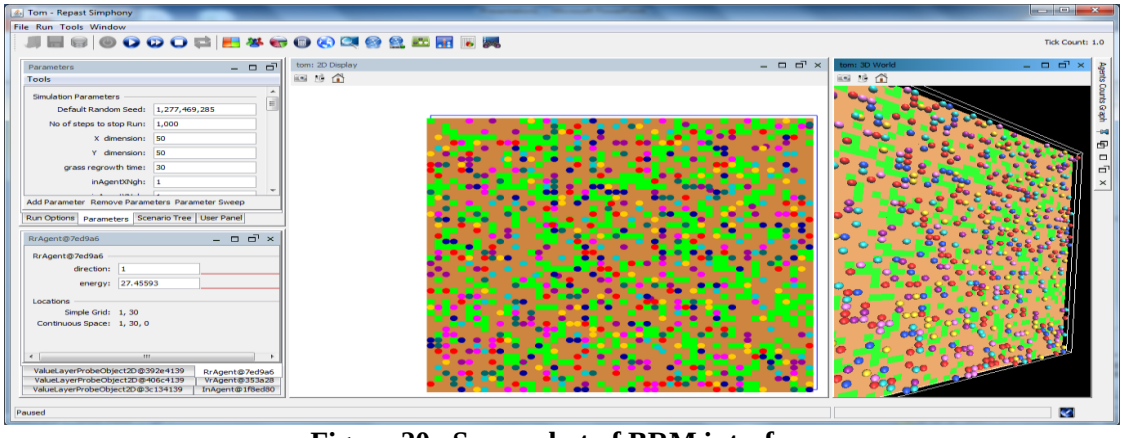


Figure 20. Screenshot of BRM interface

2.2.6 BRM hypotheses and predictions

There are two main agents in BRM; agents which are capable of understanding others' beliefs regarding the location of the food and agents with no belief representation ability. The dependent variable is the level of belief representation and the independent variable is the performance of agents in a competitive environment.

The hypothesis is that agents' with belief representation ability perform significantly better than agents without this ability in the simulations. Also, the prediction is that agents' performances in consuming food is related to their ability to understand others' beliefs. Thus, the expectation is that belief representation ability acts as an effective factor in agents' performances.

Agents' performances are determined by the total number of the food they have consumed at the end of each simulation run. The average performance is calculated by running the simulation for four times. The main analysis strategy in BRM is to compare each agents' performance within different parameters of the virtual world. Further criteria to compare the agents' efficiency include the number of false beliefs of agents and the number of times that a function is used to create a false belief situation.

Statistical methods are usually required if results are very noisy and effects are not clear-cut. For example, if results from different parameter settings are overlapping and are not reliable. However, in BRM the parameter settings have been systematically changed and have provided clear cut results. In addition, the theoretical interpretation of the simulations focused on clear results. Hence, the application of statistical methods is not required when interpreting BRM simulation results.

2.3 The BRM simulation Results

The simulation runs consists of setups as follow:

1) Infant agent and Monkey agent setup

2) Control agent and Monkey agent setup

The results are based on the average value of 4 times running the simulation for each of the parameters. Table 2 shows the chosen parameters:

Parameter	Values
Food Number	500, 600, 700, 800
Agent Number	400, 500, 600, 700

Table 2. Parameters' values for BRM

2.3.1 Why were these parameter values chosen?

One critical reason for selecting these values for the parameters is that the number of Monkey agents' false belief is high enough to provide results which show more precise and effective outcomes of the simulation. In fact, if the number of false beliefs of Monkey agents is very low, the study cannot achieve the aim set out earlier. Thus, it is necessary first to create enough false beliefs scenarios.

Subsequently, it is possible to evaluate the agents' performance in the context of false beliefs. Therefore, in situations where Monkey agents' false beliefs do not happen there is no point in running the simulation to find the effect of false belief understanding. To illustrate this point, it is necessary to explain when this situation might happen.

The main preconditions for creating false beliefs scenarios for Monkey agents are:

i) There is at least one registered food in their memory from the past time step.

ii) The previously seen food no longer exists.

iii) There is no food available in their field of view.

iv) It is possible to move to the previous cell where the food was seen.

These criteria demonstrate that the number of food and the number of each type of agents in the environment are the key parameters. Furthermore, the ratio of number of food to number of agents has a direct effect in meeting these conditions, which will be explained through this section.

Another reason for selecting these values for parameters is to avoid the extreme values, which make the environment unpredictable and uncertain. For example, when the number of food is less than 300 the uncertainty of food distribution is high. In contrast, when the number of food is high, agents are able to consume food without applying their plans. In other words, because the number of food is excessive in the environment, agents readily consume food without considering their rules and thus the perspective of other agents is not relevant any more. For example, Infant agents' ability regarding Monkey agent is not relevant to their performances. The chance that false belief scenarios happen decreases sharply in the case that the number of agents is less than 300, especially when the number of food is low as well. Moreover, when the number of agents is more than 1000, this means that more than 2000 agents compete in a dense environment (due to two types of agents participation simultaneously). The lack of sufficient freedom to move is another obstacle for agents to apply their plans. Thus, the selected values for parameters create optimum situations for the false belief scenarios to happen in BRM.

2.3.2 Agents' performances

One of the main objectives of BRM is to explore the effects of parameters on agents' performances. Conceivably, the number of consumed food by agents is a measurement to evaluate the agents' performances. Infant agents reflect the ability to understand others' beliefs whereas Monkey agents are only able to remember the location of food. Moreover, Control agents are able to track others' beliefs. Therefore, by comparing the agents'

performances, it is possible to identify some patterns and links between agents' performances, their abilities and the concepts behind their abilities.

The simulation results illuminate the answers to many central questions: Which type of agents performs the most effective in this dynamic world? Why do these types of agents' perform better? What factors improve the agents' performances? Are there any critical preconditions to their high performance? If so, what are these pre-conditions? The following graphs which illustrate the differences between Monkey and Infant agents' performances, and also the graphs which show Monkey and Control agents' food consumption address these questions.

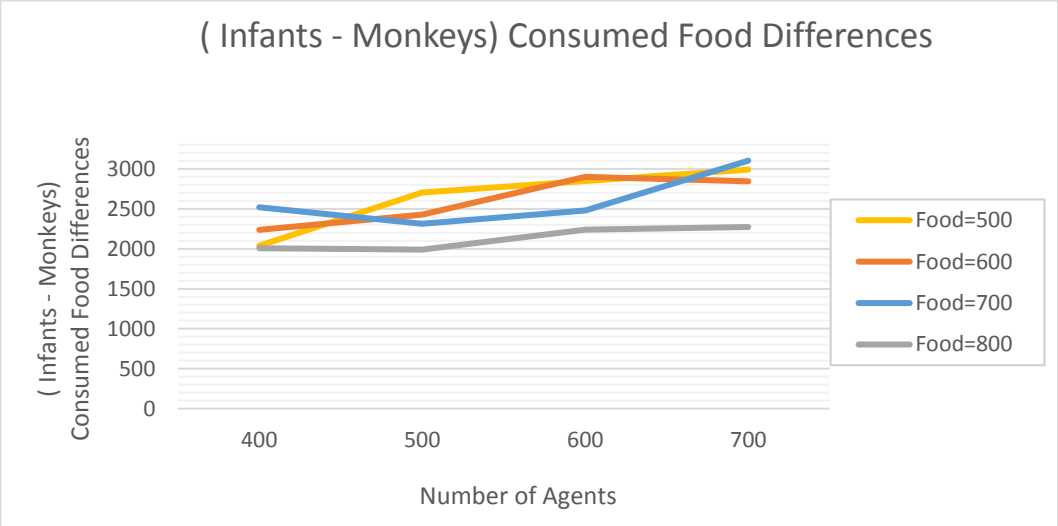


Figure 21. The difference between Infant agents and Monkey agents

The general performance of Infant agents and Monkey agents' differences is illustrated in Figure 21. The only consistent pattern for performance differences occurs when the number of food is equal to 800 where the performance differences show the lowest variation and lowest value. The difference between Infant agents and Monkey agents' performances ranges from 2000 to 3100. The most salient performance difference happens when the number of agents is equal to 700 for each type of agents; nevertheless primarily, it is subject to the number of food. For example, when the number of food is 800, the differences decrease. One main reason is that the high availability of food makes the agents

consume food without the occurrence of false belief situations. The prominent pattern shows that performances differences increase as the number of agents' increases. However, the number of food has a great impact on this pattern. In fact, the lack of a solid pattern indicates the complexity and dynamics of the environment which is dependent on the fraction and interrelationship of the parameters.

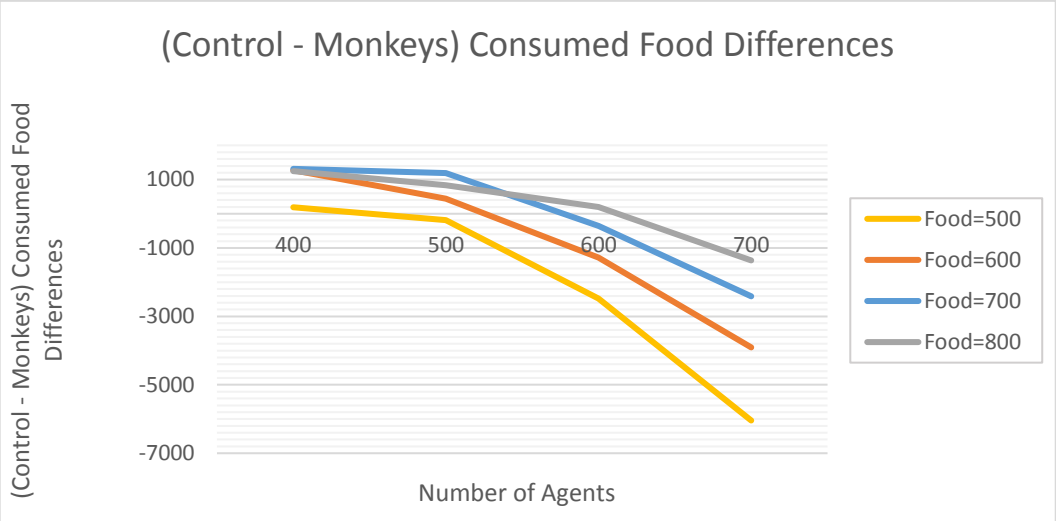


Figure 22. (Control Agents – Monkey Agents) Consumed Food

Note that a simulation setup with Infant agents and Control agents yields no meaningful results. The reason for this is that Infant agents are designed to understand others' false beliefs while Control agents have no memory to store the information and follow the previous food, thus, the preconditions for Infant agents to follow their strategy in regards to understanding others' belief is not met by Control agents. Therefore, a simulation setup with Infant agents and Control agents is misleading.

However, Figure 22 shows the difference between Monkey agents and Control agents' performances and by making a comparison between this figure and Figure 21, it is possible to compare the performances of Infant agents with Control agents. Figure 22 shows that as the number of agents increases, the performances differences increase. Whereas, with an

increase in the number of food the difference decreases. The performance of Control agent is higher than Monkey agents only when the number of agents is equal or less than 500 and the number of food is higher than 700 as Figure 22 shows. Nevertheless, Monkey agents perform more effectively than Control agents, when the population of the world is higher than 1000, particularly when food availability is low for example 500. Thus, the performance of Monkey agents is consistently higher than Control agents when the number of agents increases and the number of food is relatively low. This contradicts the initial expectation that the wider field of view should improve performance in general. In fact, in this situation they are using their strategies more often. Thus, as they are using their rules, the field of view has no effect on their performances. In addition, Figure 22 shows that as the number of food increases for constant number of agents, the difference between Control agents and Monkey agents starts to decrease; because agents simply start to consume food with less effort rather than applying their distinct rules. However, in the case that there is no food available, Monkey agents can still apply their strategies, remembering the food from past time step, whereas this does not apply to Control agents.

Noticeably, by comparing Figure 21 and Figure 22, it becomes clear that the differences between Infant agents and Monkey agents are consistently higher than the differences between Control and Monkey agents' performances in both of the above situations. For example, in the case that Control agents perform better than Monkey agents, the number of agents are 400 and number of food as high as 800, which indicates that agents can achieve food easily; the difference between Control and Monkey agents is half of the difference between Infant and Monkey agents. This clarifies that Infant agents always perform more effectively than Control agents. These results indicate that Infant agents demonstrate the most efficient performance in the simulation. The Infant agents' ability to consider others' beliefs is their key to success.

The differences of Monkey agents' consumed food in simulation with Control and Infant agents are shown in Figure 23. It demonstrates that Monkey agents' performance in competition with Infants is less effective than the competition with Control agents. In other words, Monkey agents perform in a lower level with Infant agents than Control agents.

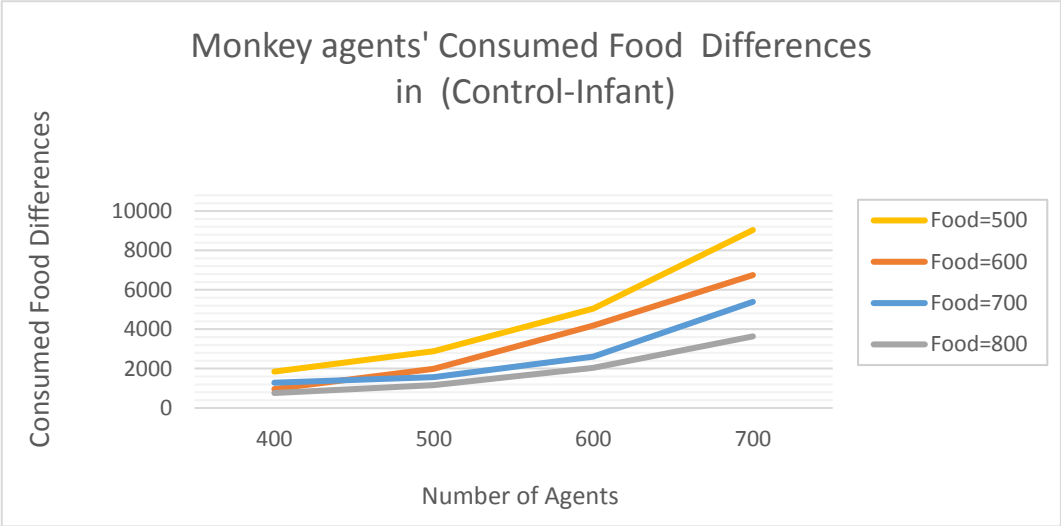


Figure 23. Monkey agents' Consumed Food Differences graph
Simulation setup of (Monkey, Infant) and (Monkey, Control).

2.3.3 The Number of false beliefs of Monkey agents

At Monkey agent's false belief scenarios, Infant agents send a message to the log window of the simulation environment as an output that there is a false belief scenario regarding the specific Monkey agent. In addition, at the end of the simulation run, there is a message displaying the total number of false beliefs that has happened to Monkey agents. More precisely, when in Monkey agent's memory, there is a food in a specific location but in reality there is no food at that location based on Infant agent's perspective, and Monkey agent has no alternative food in its neighbourhood, then the number of Monkey agents' false beliefs increases by one. Note that in each time step, when there is more than one false belief, for each Monkey agent, only one is considered. In order to calculate the number of Monkey agents' false beliefs in the simulation set up of Monkey and Control agents,

because Control agents have no memory about Monkey agents perspective, it is necessary for the Control agents to temporarily have access to that memory to check if Monkey agents had a false belief or not. Afterwards, this access ceases and they cannot use it for any other purposes. The number of Monkey agents' false beliefs in both simulations, (Infant, Monkey) and (Control, Monkey), is illustrated in Figure 24.

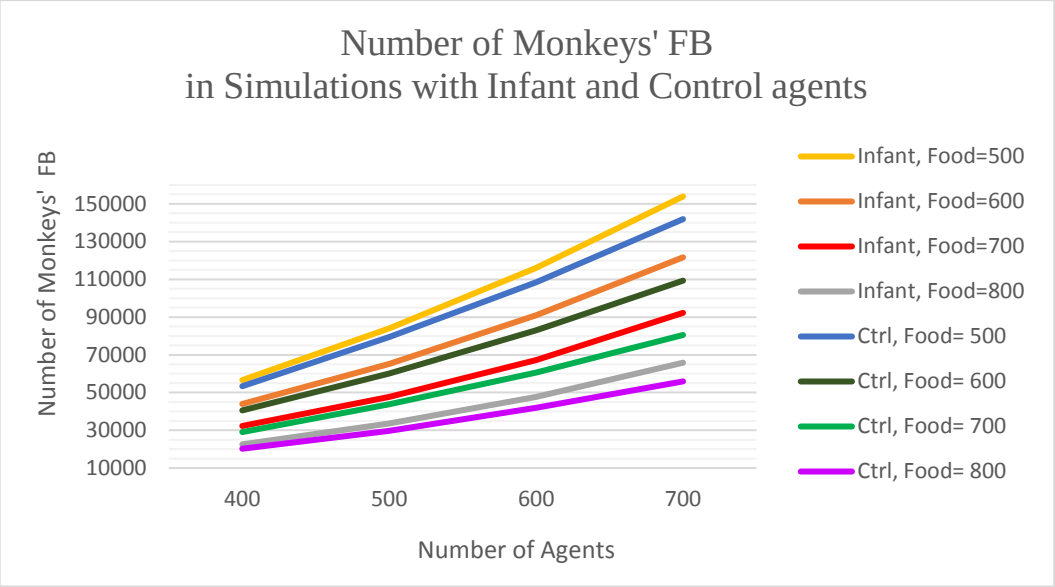


Figure 24. The number of false beliefs that happens to Monkey agents
Simulation with Infant and Control agents, Food=500, 600,700, 800.

The number of Monkey agents' false beliefs starts to increase as the number of agents increases, in simulation setup with Infant agents and with control agents. The reason is that more Monkey agents are able to register the food and pursue it in the next step. Figure 24 demonstrates that with the availability of 500 to 800 food items, the number of Monkey agents' false beliefs is higher in competition with Infant agents than in competition with Control agents. Consistent with the previous results, the graph indicates that Infant agents create more false belief situations for Monkey agents. This indicates that the aim of creating false beliefs, which is set by Infant agents, is successful. Also, Figure 24 shows that Infant agents perform more effectively than Control agents.

In general, Figure 24 shows that the number of Monkey agents' false beliefs is negatively correlated with the number of food but it is positively correlated with the number of agents. As the number of food increases, the number of false beliefs decreases. Conversely, as the number of agents increases, the number of false beliefs increases. Similarly, Figure 24 demonstrates that as the number of food increases to 800, the number of Monkeys' false beliefs declines in simulations with Infant and with Control agents. Intriguingly, the line graphs of both simulations converge gradually as the number of agents decreases. Likewise, as the number of food increases to 800, this convergence gradually develops and at the same time the number of Monkey agents' false beliefs decreases. Thus, the ratio between the number of agents and the number of food has a direct impact on this variation.

2.3.4 Priority Function

Infant agents use the priority function, which has been explained in Infant agents' strategy, when there is more than one food in the field of movement. By applying the priority function, Infant agents prioritise moving towards the food that might create false beliefs for monkey agents. As the number of agents increases, the number of times that the priority function is used increases. However, by increasing the number of food (except for 600), the number of times that the priority function is used decreases as is shown in Figure 25.

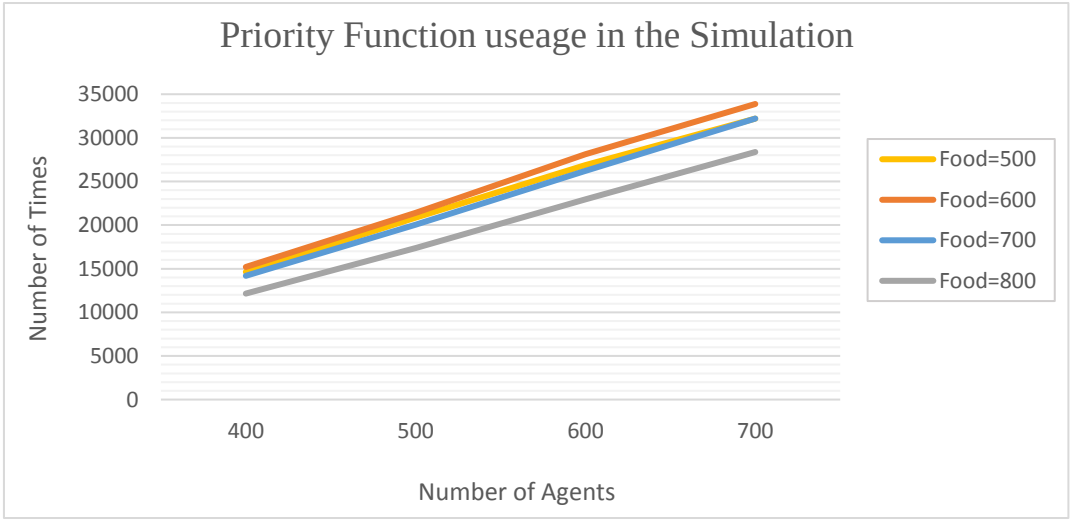


Figure 25. The graph of number of times that Infant agents used the Priority Function

For the number of 600 food, applying the priority function is optimal; nevertheless, as the number of food increases to 800, agents apply the function less because the rate of food is high enough to consume it, without using their strategies. In these situations, the number of false belief scenarios becomes lower and therefore the differences between agents decreases.

2.4 In what way is BRM an effective model?

This thesis argues that BRM is an effective model corresponding to the standard false belief task. BRM is a useful model in that its results clarify the concept of belief reasoning and it is able to answer some of the questions posed in the introduction. The model accuracy evaluation is possible through its validation and verification. Verification involves processes which verify whether the computational model follows its specifications and planned functions. Whereas the validation process assesses the computational model and its code and how they represent the aim of the model. Moreover, sensitivity analysis is useful to examine when the results are dependent on the model's parameters. Sensitivity analysis is possible by changing the computational model's parameters systematically to check whether the simulation results are changing. The explanation of BRM validation and verification processes provide an evaluation for model accuracy as follows:

2.4.1 BRM Verification

The procedure of understanding how a model works advances our understanding of the reasons of the proceedings in the model (Wilensky & Rand, 2015). BRM verification is discussed in the methodology section 2.2 by explaining the conceptual plan of the model and its arrow and box diagrams that depict the flow of the reasoning in false belief scenario of the model. For example, Figure 14 illustrates an activity diagram of Infant agents' priority function. In addition, Table 1 clarifies the abilities of different type of agents'

regarding tracking others' field of view and registering their own or others' perspectives and Figure 18 shows Infant and Monkey agents' perspectives about the location of food in the past and present time steps. All of these conceptual descriptions of Infant and Monkey agents determine how they are carrying out the intended false belief setup. Besides, the details of how false beliefs scenarios occur and how infant agents recognise them is already explained in 2.2.4 section.

Furthermore, BRM sensitivity analysis tests are explored by altering the model's initial conditions; two parameters of the initial setting, the number of food and the number of agents, have been altered to various values to examine the model sensitivity to its initial conditions. The BRM results show that agents' performances are sensitive in uncertain environments where it is not possible for agents to apply their rules and strategies. The ratio between the number of available food and the number of agents is one of the factors that determines how much food is enough for agents. In the case that the number of food is high enough, agents can consume food without using their strategies, thus agents do not use their specific strategies. The other related factors include the proportion of the number of food to the number of cells in the environment and the proportion of the number of agents to the number of cells in the environment. In respect to the verification testing and code check, for each sub function a small programming test has been conducted to examine whether its programming code works correctly.

2.4.2 BRM Validations

In terms of BRM validations, a detailed comparison between the model and the standard false belief task is described later in this chapter. By this analogy, the micro level rules of false belief procedure in the model have been elaborated which demonstrates the corresponding relationship between BRM and the practical false belief task. Moreover, the BRM's macro level validity has been shown for each parameter through the simulation

results and its graphs. The aggregated results enable BRM to conclude general statements that practically match with the false belief task; this has been already explained in the section 2.3.2.

Throughout the BRM implementation, the design and programing of agents have been constantly tested to check whether it theoretically represents the false belief concept and if not the design has been modified subsequently. This modification cycle had been continued to ensure that the preceding model truly corresponds to the false belief process.

2.4.3 The analogy between the standard false belief task and BRM (Validation)

Sally and Ann false belief task version, as already described, is used in BRM and the fit between the BRM and the false belief task is demonstrated in Table 3 by comparing their corresponding critical features. More specifically, Infant agents are analogous to the participant child as they are active and moving towards the food, while the Monkey agents play a role similar to that of Sally. Any agent, which consumes the food that Monkey registered earlier, acts as 'Ann' in the task. Similar to Sally's registration regarding the location of the ball in the basket, the Monkey agent registers the location of the available food in its field of view to use in the next time step. When Sally leaves the room it is similar to when the Monkey agent moves, causing the food to be no longer in its field of view, both unintentionally create environments, which have the potential for false belief scenario. Infant agents perceive Monkey agents' field of view which is shared with their own. Notably, Infant agent is only able to store each Monkey agent's perspective as long as it exists in Infant agent's field of view. Moreover, similar to the child who is capable to pass the Sally and Ann false belief task, Infant agent is able to recognise the Monkey agent's perspective and predict its desire towards a registered food in its memory. This prediction associates with the Infant agent's priority towards the same food as the Monkey agent,

which signifies Infant competency of passing the false belief, and is analogous to the child ability to pass the false belief task.

Accordingly, the perspective differences between Infant and Monkey agents are related to the location of food. The real location of the food is not the same as the Monkey agent's perspective because the Monkey agent is unable to update the current location of the food. In contrast, the Infant agent has access to the real information as well as the Monkey agent's perspective, both of which provide key information for the false belief task.

Hence, the BRM represents the belief attribution to Infant agents similar to the child in the standard false belief task. However, these belief attributions occur simultaneously for a number of Infant agents relating to Monkey agents' beliefs as they interact within the environment. These interactions between agents create a number of typical false belief tasks at the same time. Importantly, the social competence aspects of the belief attribution start to emerge naturally by the dynamics of this virtual society. A variety of own and others' true and false belief scenarios develop through the simulation far beyond the isolated version of false belief task.

Sally → Monkey agent, The child → Infant agent, Ann → Other agents, Ball → Food, Basket → Cell, Room → Field of view	
Sally and Ann False Belief Task	BRM
Sally registers the location of the ball in the basket.	Monkey agent registers the location of the food in a cell.
Sally leaves the room.	Monkey agent moves to another cell and can no longer see that food (out of its field of view).
Ann moves the ball to her box.	An agent consumes the food (which Monkey agent has in its memory from the last time step).
Sally returns to look for her Ball.	Monkey might return to look for the food.
The child is asked where Sally will look for the ball.	Infant will recognise that if Monkey agent returns, it moves towards the food.

Table 3. A comparison between Sally and Ann false belief task and BRM

2.5 DISCUSSION

2.5.1 Belief Representation effects on Infant agents' performance

The Belief Representation Model (BRM) consists of three agents:

- Infant agents are able to pass false belief task.
- Monkey agents are able to remember and track the food from past time step.
- Control agents are able to track others' field of view regarding the food. They are used as a control measurement.

The results of the simulation between Infant agents and Monkey agents' consumed food differences is gathered and show that Infant agents consistently perform better throughout the different values for the parameters including number of food and number of agents.

Infant agents recognise Monkey agents' belief representation; they track Monkey agents' field of view, register and store Monkey agents' perspective regarding the food. Moreover, BRM involves reasoning of beliefs and desires of agents' behaviour. Infant agents reason about Monkey agents' desires and beliefs. They also inhibit their own perspective regarding the location of the food and apply Monkey agents' perspective.

In BRM, all agents' goal and desire are to move towards the food and consume it. More than one agent may have the desire to consume the same food. In the case where there is only one food to select, it becomes their only choice on that time step. Once an agent has access to more than one food, its reasoning style is to select one food from several, and it is influenced by its different capabilities and the information it perceives from the environment. For example, Infant agents' reasoning style rests on their competence of understanding Monkey agents' false beliefs. Infant agents' preferences are for the food which effect Monkeys' true beliefs. Each Infant agent's representation is contingent on its perception, the information in its memory, and the way that it understands other agents'

beliefs and the environment. In fact, Infant agent's memory stores some experiences of Monkeys agents' perspective.

At this point, Infant agents are able to recognise Monkey agents' false beliefs. However, this recognition is not sufficient to perform successfully, it still requires using the information about this understanding together with other online information to set an effective plan, and subsequently to take an action based on the plan, to perform more efficiently. In other words, the efficiency of Infant agents' performance is because of a series of proceedings: Firstly, their recognition of Monkey agents' true and false beliefs regarding the location of the food and the related information of their field of view. Secondly, applying this understanding and information into a plan or a rule that enhances their chance to achieve their goals. Thirdly, performing an action by employing the plan. For example, in BRM, Infant agents create false belief scenarios for Monkey agents; subsequently Infant agents consume more food by increasing false belief situations for Monkey agents.

There is no doubt that a combination of understanding others' beliefs, reasoning about this information, planning and contributing an action based on this, are the main factors in producing a more efficient performance for Infant agents.

2.5.2 Differences between Monkey agents and Control agents

The main ability of Monkey agents includes storing the information about the location of the food and using it in the next time step when there is lack of food. Although, they remember the location of the food from the past time step, they are egocentric and do not consider other agents' perspective. In contrast, Control agents are able to track other agents' perspective regarding the food in the present time step. However, they have no memory and do not remember others' beliefs from the past. The BRM results demonstrate that Control agents' performance is less efficient in comparison with Monkey agents,

particularly when they are using their own strategies rather than the situations in which there is enough food in the environment to consume without using their strategies. However, in situations where the number of food is more than half of the number of agents, Monkey agents mostly do not use their strategy whereas Control agents' strategy is still applicable. Therefore, in this situation Control agents' performance is slightly higher than Monkey agents. Whereas, the Infant agents' performance is more than twice as high as the Control agents' performance in the same set-up with Monkey agents. These results signify that agents' abilities and strategies in micro level can influence their performances in macro level when their strategies are put into practice.

2.5.3 Infant Agents' diagram (IAF)

One important approach in agent-based models is to present the flow of data in a diagram to show the sequence of operations and processes performed within the system. A diagram that illustrates the control flow of the agents' actions can represent the underlying logic of the complicated and interconnected procedures of the actions. Infant Agents arrow and box diagram, which is illustrated in Figure 15, represents the underlying basic phases that occur for an agent with understanding of others' false beliefs. Thus, this section explores the concept behind each phase of this diagram and investigates the false belief procedure in a structured and coherent approach by classifying the procedure in to four phases as follow:

2.5.3.1 - Collecting Information

Collecting information is a central phase in BRM; Infant agents collect essential information from their field of view in every time step. They collect the information about the location of food, the location of other agents and specifically the information about the location of food from the Monkey agent's perspective. Infant agents reason about which information they need to collect. For example, they are interested in Monkey agents'

perspective, as they both have access to the same food. In comparison with Sally and Ann false belief task questions, the BRM questions for Infant include:

- Where does the Monkey agent store the location of food (which cell)?
- Is the food still in the Monkey agent's field of view?
- Can the Infant agent consume the food, which is stored by Monkey agent? (Has the food been eaten by other agents?)
- Where will the Monkey agent search for the food, when it returns?

These questions can be answered through the IAF phases. Infant agents collect the necessary information from their field of view, which are particularly related to the Monkey agents' perspective of the location of the food. There is a dynamic link between the collecting information phase and the other phases of IAF in regards to each false belief scenario in BRM. The collecting information phase is parallel to the time and dynamics of the world; meaning that the collecting information phase is a continuous process corresponding to the time steps and environmental changes. These changes including agents' movements and the location of the food create a dynamic world. Agents collect information in every time step in this dynamic world. In other words, agents continue to collect others' belief scenarios' information as the world changes through the movement of agents over time. Thus, the online raw information becomes available from the collecting information, which can then feed other phases of IAF to complete their related processing simultaneously.

2.5.3.2 - Recording Information

In order for Infant agents to succeed in understanding others' false beliefs, BRM demonstrates that there are memory demands on Infant agents to store data from the Monkey agents' perspectives about the location of the food. BRM highlights the role of

memory to store Monkey agents' perspective information. In the absence of memory, it is not possible for the Infant agents to retrieve this information in the next phases.

2.5.3.3 - Reasoning Process of Beliefs and Desires

This phase of IAF involves complex information processing and demonstrates the capability of the Infant agents in understanding others' false beliefs. Conceivably, there are two different versions of beliefs about the location of the food in false belief scenarios; the Infant agents' own perspective, which is the last updated version of the reality, and the Monkey agents' perspective which is not updated from the last time step. By default, agents use the last updated information about the location of the food due to the dynamics of the environment. However, firstly in this phase Infant agents inhibit their regular information temporarily and restore the Monkey agents' beliefs. In other words, similar to the Sally and Ann task, the Infant agent inhibits its own belief about the current location of the food and retrieves the Monkey agents' perspective which has already been stored.

Secondly, the Infant agents reason about the Monkey agents' beliefs and desires; Infant agents take into consideration that other agents have a common desire towards the food. The agents' common desire causes a competitive environment. Subsequently, because Infant agents track others' field of view and store others' belief about the location of the food, they reason that if other agents have the same beliefs about the location of food, they are their competitors. They recognise Monkey agents' beliefs, including true and false beliefs. The processes that Infant agents understand Monkey agents' false belief has already been described in methodology. However, a brief explanation clarifies the reasoning phase. The procedure of Infant agents' understanding Monkey agents' false belief includes:

- Infant agents inhibit their own belief about the location of food temporarily.

- Infant agents retrieve the information of Monkey agents' perspective about the location of the food.
- Infant agents consider that there is food in location X in terms of Monkey agents' perspective, when there is no food in location X in reality. In addition, at the present time step, there is no alternative food for Monkey agent from Infant agent perspective. In other words, if there is an alternative food for the Monkey agent (from Infant agent's perspective), creating a false belief does not really affect Monkey agent's performance.

In sum, Infant agents reason about Monkey agents beliefs by first considering its own beliefs and the common desire with Monkey agents.

Exploring this more systematically, the reasoning processes of beliefs and desires phase include three subroutines:

- Self-perspective inhibition
- Retrieving the protagonist's perspective data (from memory)
- Selective processing of protagonist's (Monkey agent) belief and desire based on its own belief and desire.

At this stage of the BRM, Infants recognition of Monkey's beliefs is complete. Although, Infant has not yet demonstrated this belief understanding in its actions but the required processes for understanding of others' beliefs is completed.

The subtle borders between understanding belief representation and exposing or using this representation are two entirely different phases of Infant agents understanding the Monkey agents' false beliefs. Analytically, based on BRM, the main procedure of belief representation finishes by the end of the third phase. Nevertheless, the last phase in which Infant agents act upon their understanding of Monkey agents' perspective remains open.

2.5.3.4 - Expressing others' mental states (actions as output)

This phase rests on the critical difference between having a competence and using it in agents' actions. In other words, understanding other agents' beliefs and desires' information is essential for the Infant agents' actions.

Noticeably, expressing others' beliefs and desires is analogous with the measurement test in the false belief tasks. Infant agents utilise the Monkey agents' belief representation in their actions. Firstly, once the Monkey agent's belief is true, the Infant agent prioritises consuming the food which creates a false belief for Monkey agents. Altogether, based on the simulation results, this strategy creates more false belief situations for the Monkey agents.

Secondly, once the Monkey agent's belief is false, then the Infant agents express their understanding of the Monkey agent's false belief by sending a message as an output that 'There is a false belief regarding the Monkey agent *number-x'*. At the end of the simulation, the message shows the total number of false beliefs that Monkey agents had and Infant agents recognised. Thus, the Monkey agents' false beliefs have no direct effect on the Infant agents' actions in BRM. However, each message signifies the understanding of Monkey agent's false belief by an Infant agent.

2.5.4 Applying minimal theory of mind principles to agents

Butterfill and Apperly (2013)'s minimal theory of mind presents the ability to automatically track others' belief-like perceptions without representing any sophisticated psychological states. Minimal theory of mind is based on four principles: goal directed action, field and encountering, successful registration and action influenced by the registration, which has been already described in the general introduction. Apperly and Butterfill suggest that an individual with minimal theory of mind ability could pass many tests that were supposed to be acid tests of theory of mind such as false belief tasks. They also state that other principles and variation on the principles can be added to provide a

wider range of theory of mind abilities (Butterfill & Apperly, 2013). These four minimal theory of mind principles are applicable to Control agents and Infant agents in the model as follow:

Principle 1: Each agent's goal is to move towards food.

Principle 2: Each agent is able to see a limited local part of the world around itself as its field of view. Agent encounters and perceives the food and other agents which are in its field of view. An agent first encounters food in its field of view before any goal-directed action is taken for the food. These principles are described in the first phase of IAF.

Principle 3: Each agent registers the location of the food in its field of view correctly. Therefore, the agent simply has a belief that there is food in that location. The relation between the agent, the other agent and the food is defined by the registration. This principle has some similarity with the second phase of IAF, which will be explained later.

Principle 4: After a successful registration, each agent, based on its beliefs about the location of the food and its desire to consume the food, detects other agents' beliefs and desires regarding the food. Subsequently, the agent takes action and moves based on its perspective of other agents' beliefs and with other information from its field of view.

Thus, agents use both other agents' beliefs and the current information about the world to make decisions and act. The fourth principle in minimal theory of mind can be considered parallel to the fourth phase in IAF (an action).

Therefore, Control agents and Infant agents have minimal theory of mind. Nevertheless, Infant agents do not use minimal theory of mind only. They are also able to infer and understand others' belief through reasoning, self-perspective inhibition and allocating memory to others' beliefs. In contrast, Monkey agents fail minimal theory of mind ability because they do not hold all of the corresponding principles. For example, Monkey agents do not track others' field of view.

2.5.5 True Beliefs in BRM

This thesis argues that IAF is also applicable for true belief representation. The main difference between true and false belief representation reflects the match or mismatch between the protagonist's perspective and the real situation. BRM includes both true and false belief scenarios for Infant agents in which a generic belief representation systematic approach can be developed.

2.5.6 The Network in IAF

IAF presents the key components of the belief representation processes, which consist of perception, memory, inhibitory control and selective process reasoning, in addition to complex reasoning resources, which is essential for expressing others' belief. Together, these components represent a network of resources that shapes the individual's ability to understand others' beliefs which is compatible with the developmental literature underpinning theory of mind network (Mohnke et al., 2015) (Gallagher & Frith , 2003) (Carrington & Bailey, 2009).

2.5.7 Complexity of the environment

In each time step, each agent interacts within its local environment resulting in more changes in the world. These dynamic changes to the world will affect the agents' decisions and actions. Beliefs based upon events in the past, may no longer be true in the present due to the dynamic nature of the model. All of the changes that happen between any two adjacent time steps are not observed by some agents, whereas they might be observed by some other agents, which might create false belief scenarios. In addition, the random placements of the food and some random movement of agents together with dynamics of agents' actions escalate the uncertainty of the social environment. The dynamics of the simulation are high enough to change the situation regularly. For example, in the event

where there is no food in a particular cell at time step t, it is possible that this cell fills with food at the next time step (t+1). This is due to the random nature of food replacement.

2.5.8 Imperfect Perception

There is a general rule in agents' vision ability that makes some of the information incomplete and imperfect; vision is more comprehensive in the centre of field of view, but as it reaches the edges of the field of view, the information is not as complete. In fact, as the shared area between field of view of two agents increases, the information they acquire from each other is more reliable and complete. This rule is simplified in Figure 26 by illustrating a square as the field of view of yellow agent. The arrows show the direction from comprehensive information in the centre towards the edges with more imperfect information. This information concerns objects or other agents in this field of view. For instance, consider the grey area as yellow agent's field of view, the red square as the red agent's field of view and the blue square is the blue agent's one. The yellow agent is able to perceive precise information about the red agent, which is closer to the centre. The information about the blue agent that is closer to the edges of the grey square is less precise for yellow agent. Since the agents' vision is depend on the size of field of view, agents' perceived information is also limited.

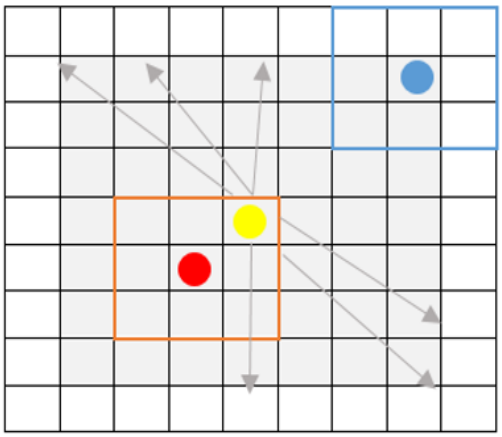


Figure 26. Incomplete Information

The arrows show the flow from comprehensive information towards incomplete information. The yellow agent has more precise information about the red agent in the centre, than the blue agent at the edges, considering the grey square as its field of view.

2.6 Conclusion

BRM is a reliable model which illustrates the underlying processes of Infant agents, which are capable of understanding others' false beliefs, in a structured and coherent approach by classifying this procedure into four phases. The first phase, is called collecting information, in which agents collect information particularly in relation to other agents' perspective on the location of the food. The second phase, recording Information, is when agents store the collected information including the other agents' perspectives in their memory. This phase highlights the role of memory in belief representation. The third phase, which is called reasoning process of beliefs and desires, is the main phase for processing the information about others' perspectives. In this phase, agents inhibit their own beliefs temporarily and restore others' beliefs. Another part of this phase involves critical reasoning about others' beliefs and desires. The last phase is concerned with deciding on an action by considering others' beliefs and desires, which is called expressing others' beliefs and desires. This phase identifies the critical difference between having belief representation ability and using it in agents' actions.

Furthermore, these four phases identify the key components of belief representation processes consisting of perception, memory, inhibitory control and selective process reasoning. These components represent a network of resources that shape the individual's ability to understand others' beliefs.

In addition, Infant agents in BRM use more than minimal theory of mind as they are able to understand others' beliefs through reasoning, self-perspective inhibition and allocating memory to others' beliefs. In contrast, Monkey agents fail the minimal theory of mind principles.

Moreover, Infant agents perform better in the environment because: Firstly, they are capable of understanding others' beliefs regarding the location of the food. Secondly, they

apply this understanding and information to a plan which enhances their efficiency in achieving their goals. Thirdly, they perform an action by employing the plan. Thus, the main factors in producing a more efficient performance for agents include a combination of understanding others' beliefs and implementing this understanding by taking action.

CHAPTER 3

3. AN AGENT-BASED MODEL TO UNDERSTAND A SIMPLE THEORY OF MIND

3.1 INTRODUCTION

Surprisingly, Mary left her wallet at home. Her mother found it. She thinks that Mary thinks that her wallet is lost.

Humans are able to infer others' mental states behind their actions. How can one construct mental states connections with others? This mainly invisible connection relies on cognitive resources.

People regularly understand others' behaviour by attributing them mental states such as beliefs and desires. They make inferences about others' unobserved mental states from the observed behaviour. These inferences are often precise and indispensable for humans' social life. In general, people's beliefs, desires, emotions, and other mental states are a dependable guide to their future actions. It is also possible to explain one's action based on her mental states. Therefore, her planning process needs to be inverted; thus, they reason backwards to infer others' beliefs and desires from their actions. In fact, inverse planning concerns working backwards from the action to the underlying mental states, to make inferences about beliefs and desires that caused one's action (Baker, Saxe, & Tenenbaum, 2009). Accordingly, psychologists apply the same approach known as belief-desire-action reasoning. One's beliefs and desires are the reasons for their goal-directed actions; this demonstrates a coherent picture of every day mental inferences (Wellman, 1990). Intriguingly, there is a body of evidence suggesting that human infants construe actions compatibly with an inverse planning paradigm (e.g. Baker, 2012; Phillips & Wellman, 2005). Furthermore, "the expectation that agents will plan approximately rationally to achieve their goals, given their beliefs about the world" (Baker, 2012, p. 33) are known as the principle of rationality. Yet, one-year infants are able to use the principle of rational action to interpret and predict goal-directed actions and to make inferences about unseen aspects of their context (Csibra et al., 2003).

This chapter of the thesis raises several fundamental questions about theory of mind processes in human minds. What are the characteristics of the levels prior to a simple theory of mind ability? How do one's own beliefs and desires interact with others' beliefs and desires mentally? What is the shared set of basic processes of different varieties of theory of mind tasks? What are the steps of reasoning in understanding others' beliefs and desires, including inverse planning? How does reasoning affect the level of complexity in theory of mind? Moreover, in terms of theory of mind advantages: How does theory of mind ability influence the agents' performances in a competitive context? What are the other conditions, if any that makes agents with theory of mind ability perform better than others? To answer these questions and similar issues in the literature, an agent-based model has been implemented with 6 types of agents having different capabilities of understanding own and others' mental states as interacting within the environment.

The model for theory of mind presented in this chapter works similar to the two phases of IAF including recording Information and expressing phases. However, the collecting information and reasoning phases are more demanding in this model than the false belief model because here, agents infer others' desires from the observed action rather than 'seeing and knowing' scenarios. Nevertheless, these two phases are also consistent with the IAF phases but they include more steps.

This chapter explores the advantages of a simple theory of mind ability in agents' performances. For example, how understanding others' beliefs and desires improves goal-directed action's efficiency. Furthermore, this model examines how agents infer others' beliefs and desires by observing their actions, how they apply inverse planning and when they use rational principles. This agent-based model is called Mental State Model (MSM).

3.2 MSM METHODOLOGY

An agent-based model is implemented to elucidate the core concepts and connections between the indirect inferences of others' mental states and their actions in a simple theory of mind. MSM comprises of interactions between six types of agents, representing different capabilities of understanding others and their own desires and beliefs. Finally, MSM evaluates the agents' performances to achieve their goals in a competitive society.

3.2.1 Environment

The environment is made of a grid space where agents act based on their rules to achieve their targets and survive. Initially, agents are randomly placed in this environment.

At the start of each run of the simulation, the set up procedure defines parameters and their default values. It creates and places the defined number of each type of agent and the number of targets in the environment. There are other parameters, which will be explored in more detail in the next section. The environment consists of a grid of 50 by 50, through which agents move contingent on their individual rules and the position of the targets. When an agent achieves a target, that target will be removed from the environment. However, in the next step a target will be generated in another random cell. The space is toroidal therefore, if agents move to one border of the grid, it appears on the opposite border.

Time Step

The time measurement (tick) is a step in the simulation which agents simultaneously perform their actions. The default number of time step for the simulation is 1000 tick.

Neighbourhoods

The neighbourhood of an agent is the squared area around it in addition to its cell while the extent of its X-axis and Y-axis' are equal. Hence, the length of neighbourhood (Ngh) is the same in both axes' direction. For example if Ngh=3 then it includes the area of 49 cells.

Field of view

Field of view is a parameter with the default value of Ngh = 2. It defines the square of 5x5 around the agent (25 cells) within this area the agent is able to perceive the environment. The parameter of field of view changes from 2 up to 6 and one of the objectives of the simulation is to analyse the behaviour of the agents' performance by these changes.

Field of movement

Field of movement consists of the agent's first neighbourhood cells; the eight cells in agent vicinity and itself that is shown in Figure 9.

Targets

Target is a goal that each agent is required to achieve and it is demonstrated by a green cell. It is necessary for agents to reach their targets to enhance their survival rate. The survival rate is a criterion for their performance measurement. The purpose of target in MSM is similar to food in BRM. However, target is used here to represent the underlying concept of mental states whereas food cannot simply convey the impression. A parameter defines the number of targets available in the environment in each time step. For example, when number of targets are 250, it sets around one target in 10 cells of the world and the entire world consists of 2500 cells.

At the start of each time step, the targets, which were not achieved by any agent, remain in the environment. The targets, which agents achieved, will be replaced randomly in the grid for the next time step. Therefore, the number of targets in the environment remains constant through the simulation time steps

Parameters

The analysis of the agents' behaviour is based on modifying the following parameters:

- Number of agents(N):

The agents' population is a critical parameter in the environment.

- Agents' field of view(Ngh):

The area that an agent is able to observe the environment. MSM investigates the effect of field of view on agents' efficiency.

- Period of staying Passive(P):

The number of time steps that agents stay Passive before they change to Active mental state. This parameter has direct influence on the time steps that agents need to achieve a target. Thus, it is a key parameter and will be analysed in MSM.

- Number of targets(T):

Undoubtedly, the number of targets is a crucial factor for agents to survive in this competitive world and the analysis of this parameter is essential in the simulation.

3.2.2 Mental States (MS)

Each agent has a mental state in every time step. There are two types of mental states:

1) **Active**: Agent needs to achieve a target.

2) **Passive**: Agent cannot achieve a target.

At the start of the simulation, the mental states of agents are Active. However, their mental states change through the simulation, which are subject to reaching a target. Unless an agent achieves a target, its mental state remains Active. The mental state of an Active agent that achieves a target changes to Passive. When the agents' mental states become Passive, it remains Passive until a parameter that serves as a counter for Passive mental state, reaches zero. The number of time steps in which an agent must stay in a Passive mode is set by a parameter called period of staying Passive (P). After an Active agent reaches a target, its mental state stays Passive for a number of time steps that parameter P defines. Every agent with an Active mental state is required to achieve a target in the next time step. However, the agents with Passive mental states are not allowed to reach a target and should move to an empty cell which does not contain a target. Figure 27, the statecharts, illustrates changes from Active to Passive states and vice versa. Each Active agent that achieves a

target changes to Passive state in any time step. In contrast, an agent with a Passive mental state becomes automatically Active after staying in a Passive state for P time steps. In other words, when the agent's period increases from 1 to P during P time steps, the agent's mental states changes from Passive to the Active state.

More precisely, it can be declared as: $t_{ACTIVE} + P >= t$

Where: t_{ACTIVE} = the last time step that the agent has been Active,

t = the current time step,

and P = The number of time steps that agents must stay in Passive state.

And Active agents that become Passive, should stay Passive for P time steps.

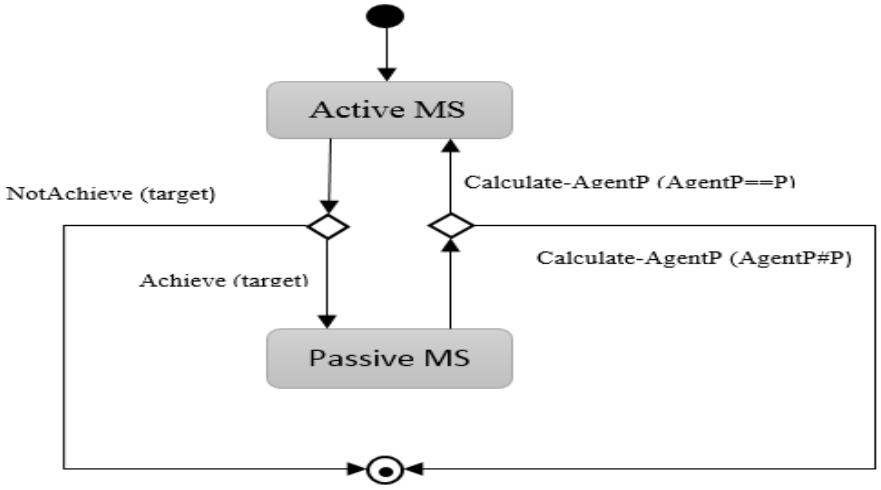


Figure 27. Statecharts of mental states changes in MSM

3.2.3 Agents' General Rules

There are six different colour-coded types of agents in the environment with their heterogeneous set of rules: Random agents (blue), Food agents (red), Control agents (grey), MinToM agents (orange), Infer agents (pink) and Reality agents (purple).

- Agents can move to a cell which is not occupied by other agents.

- If there is no option to move to another cell, then agents will stay in the same cell.

- When a target is placed randomly in the same cell of an Active agent and there are no other targets in its field of view, then it can achieve the target.

- Passive agents move towards an empty cell, which is in the neighbourhood of a target.

- Agents are able to move towards a target, which is in their field of view but is further than the immediate vicinity. In order to move efficiently, agents use an intersection set between their field of movement and the neighbourhood of the target. For example, Figure 28 shows the intersection sets (red cells) between the agent (purple circle) and the two different targets (green cross). Therefore, the agent moves to one of the intersection cells to reach its target.

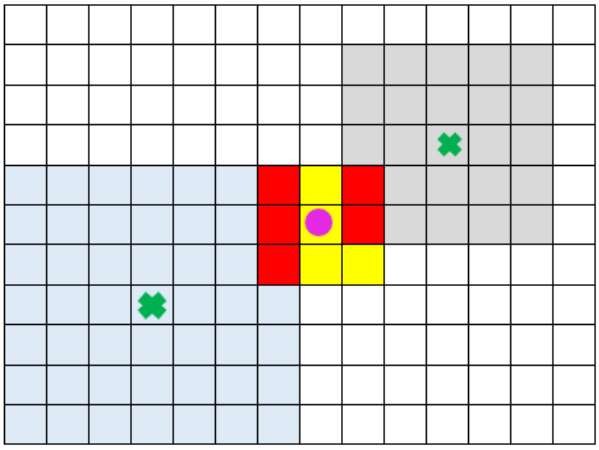


Figure 28. The intersection sets

The intersection sets (red cells) show the most efficient path between agent's (purple circle) field of movement and the two different targets' (green cross) neighbourhoods. The agent moves to one of the two red cells in the right to reach the nearest target.

3.2.4 Agents' Strategies

Agents move towards a cell depending on their mental states and their strategies in each time step. Each type of agent applies different rules and strategies contingent on their abilities. As their abilities in relation to understanding others' beliefs and desires become more effective the agents' plans and rules improve.

3.2.4.1 Random Agents' (Rm) Strategy

These agents demonstrate random behaviour, they move randomly without any specific strategy, plan or even attention to goals. Although they follow the environment rules, they have no attention to their own or others' mental states. They lack theory of mind ability in any kind and act as a control measurement.

3.2.4.2 Food Agents' (Food) Strategy

Food agents are simple reactive agents which observe their field of view and act based on the information they perceive from the environment. Food agents recognise targets and the need to achieve them. However, they are not capable of planning or having efficient strategies for the future. Active Food agents move towards the nearest cell that contains a target. In situations where there is more than one target, agents choose one randomly. Moreover, they move randomly when their mental state is Passive. They only consider their own mental states and pay no attention to others' mental states. They have no ability of theory of mind and they are used for control measurement purposes.

Food agents' perceptions, functions and memory abilities are shown in Table 4. As the table shows the Food agents' perceptions abilities include target sensor; they collect information about the targets in their field of view and use the information about the targets at the current time step. Regarding memory, they are competent in using sensory memory to keep the location of the target in the current time step. Moreover, the table shows that they only consider their own mental states.

3.2.4.3 Control Agents' (Control) Strategy

The Active Control agents move towards the nearest cell which contains a target similar to the Food agent. However, the Passive Control agents move to a cell which do not contain a target but there is a target in the vicinity of the cell. The difference between Control agents and Food agents appears when their mental state is Passive. Table 4 shows that Control agents' abilities include target sensor, sensor memory and considering their own mental

states. However, they are able to extend their field of view to search for a target if it is necessary. Figure 29 shows the arrow and box diagram of the Control agents.

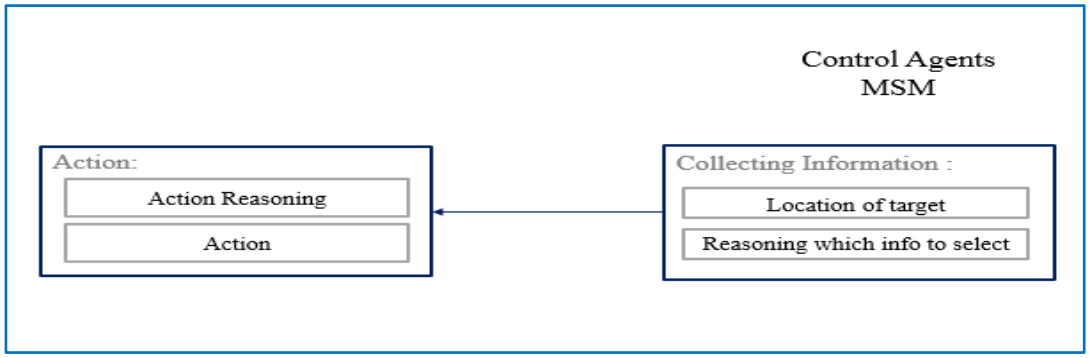


Figure 29. Control agents' arrow and box diagram

Agent's Type	Perception			Internal Functions			Memory			MS INFO
	Target Sensor	Agent Sensor	Vision To larger area	Tracking others' field	Consider owns MS	Consider others' MS	Sensor Memory	Short-term Memory	Long-term Memory	Direct access
Random										
Food	✓				✓		✓			
Control	✓		✓		✓		✓			
MinToM	✓	✓	✓	✓	✓		✓	✓		
Infer	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Reality	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 4. The agents' perceptions, functions and memory capabilities
It is based on understanding their own and others' mental states (MS).

3.2.4.4 Minimal Theory of Mind Agents' (MinToM) Strategy

MinToM agents are reactive agents which collect information about the target and other agents. MinToM agents register the location of the target from others' perspective to track their field of view in the current time step. Subsequently, they move to a cell and use this information. The general diagram of the MinToM agents including collecting information, registering others' perspective and Action is illustrated in Figure 30.

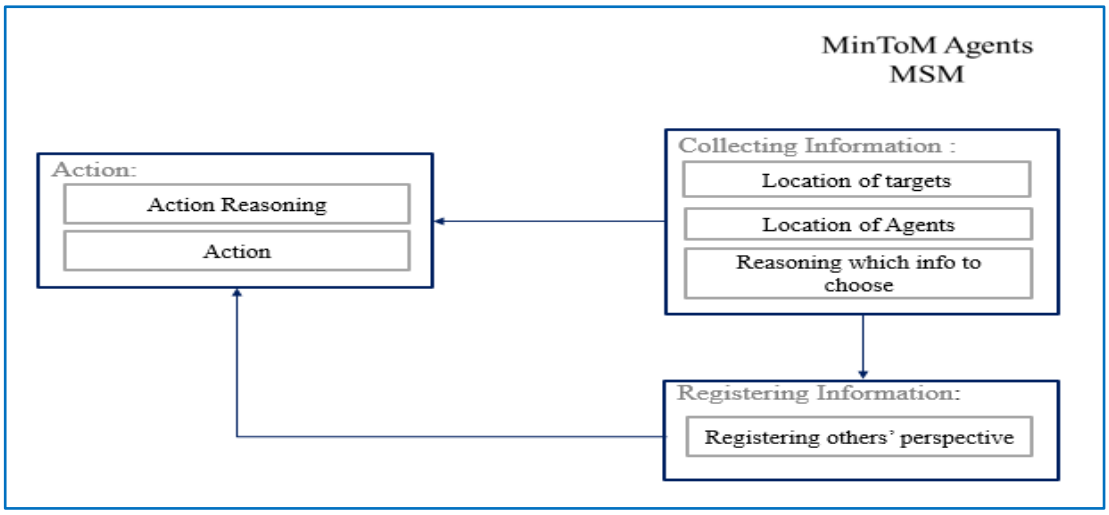


Figure 30. MinToM agents' arrow and box diagram

In general, MinToM agents move on the basis of two factors, the availability of targets and their mental states. They search for a target first in their field of movement and then in their field of view, and depending on their mental states they move to a cell with or without target. The search starts from the nearest field of view and continues to the larger area. Thus, there are various conditions in which MinToM agents apply different sub-strategies based on the situation.

In the case where there is no target in the field of movement, Active MinToM agents move towards a target with the minimum number of agents around it when there is more than one target in its field of view. On the other hand, Passive MinToM agents move towards an empty cell in the vicinity of a target and surrounded by the minimum number of competitors. In the case where they have the same number of competitors at the same distance, it chooses randomly.

MinToM agents have all of the abilities of Control agents as shown in Table 4. Their abilities are more advanced than the Control agents due to tracking others' field of view and applying short-term memory. Short-term memory facilitates agents' calculations and keeps the information for processing in each time step.

3.2.4.6 Infer Agents' (Infer) Strategy

Infer agents collect the information regarding targets, other agents, and their actions. They are able to infer others' mental states and use this understanding in their actions. Figure 31 illustrates the diagram of the Infer agents which shows they collect information, and they record, reason and act based on the information about others' perspective which will be elaborated later in the discussion.

Infer agents observe other agents' actions and infer their mental states from their actions. Principally, when they see that an agent achieves a target, they infer that the agent's mental state becomes Passive. Otherwise, they assume that the agent is Active. Suppose an agent, for example agent Y, is observed by the Infer agent as it achieves a target then:

Infer agent observes agent Y achieves a target. ====> Mental states of agent Y is Passive.

By applying the above proposition, Infer agents are able to infer other agents' mental states and plan to achieve a target. Thus, the inferences of Infer agents are based on what they perceive from the environment. Infer agents are able to remember the agent Y's inferred mental states, and use this until agent Y moves out of the Infer agent's field of view.

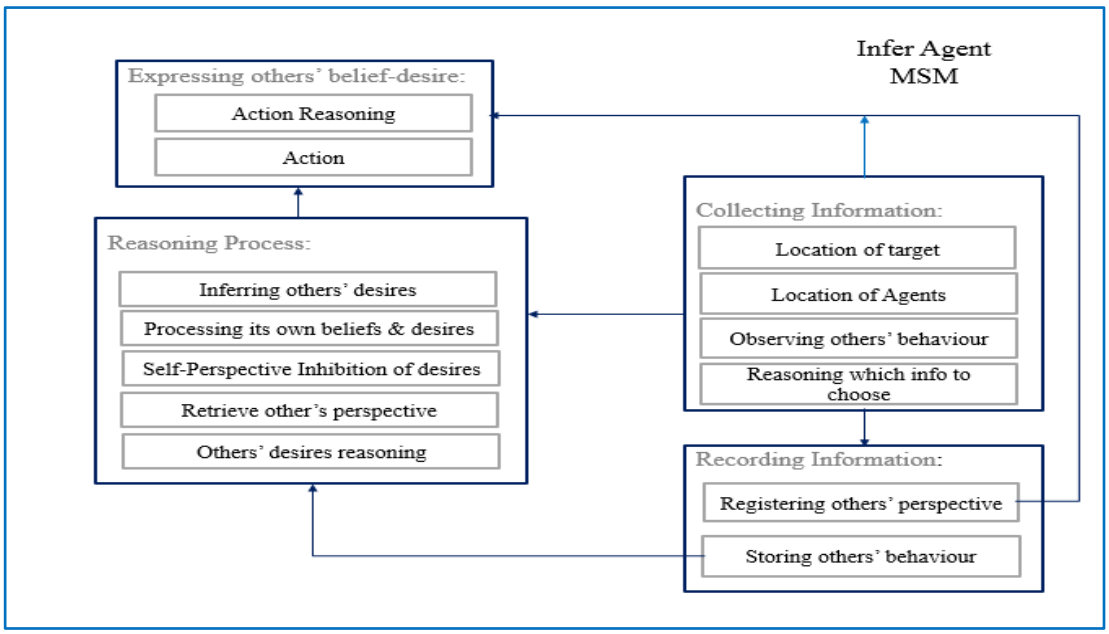


Figure 31. Infer agents' arrow and box diagram (RAF)

Working backwards from the action that agent Y has performed, the Infant agent is able to infer its desire, which caused agent Y to perform that action. To elaborate on the above proposition; when agent Y has achieved a target, it means that agent Y was in an Active states (because only Active agents are able to achieve a target) and has now become Passive (because of the rule that when an Active agent achieves a target, it automatically becomes Passive). In fact, Infant agents tacitly use inverse planning to infer others' desires from their actions by applying these inferences.

Subsequently, Infer agents employ belief-desire-action reasoning approach to agent Y by using agent Y's beliefs and desires towards the target to infer its action. Thus, Infant agent makes different decisions based on agents Y's belief and desires. Figure 33 illustrates this inference schema of Infer agent based on belief, desire and action. Therefore, Infer agents' strategy depends not only on its own mental states, but also on other agents' mental states. Infer agents with the ability of inferring others' mental states are an evolved version of MinToM agents. Thus, Infer agents' strategies are similar to MinToM agents' strategies when they are unable to have appropriate information to infer others' mental states.

In general, Infer agents' actions are based on the conditions of its own mental states, others' mental states, and the number of available targets in their field of movement and their field of view. Similar to MinToM, the search for a target starts from the nearest area to its own cell and continues to a larger area in its field of view as shown in Figure 32.

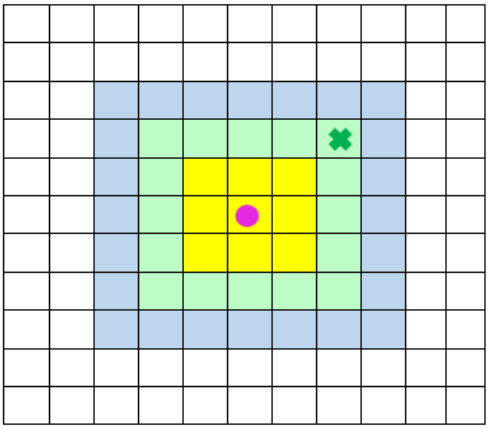


Figure 32. The search order for a target

It starts from its first neighbourhood (yellow cells) to the wider area (turquoise cells) and ends in the blue cells.

In the case where there is at least one target in the field of movement, Active Infer agents move towards the target with the maximum number of other agents (in Active state) around it. In the case where there is no target in Infer agent's field of view, then the Infer agent moves randomly to an empty cell. If there is no target in the field of movement, the action of Infer agents is subject to the number of targets in its field of view.

In the case where there is no target in the field of movement and at least one target in the Infer agent's field of view, then an Active Infer agent moves towards a target with the minimum total number of Active agents. Note, Passive agents, which can become Active when the Infer agent reaches the target should be considered within the above calculation. To explain this in more detail, suppose:

d = the minimum distance between the target and the other agent where distance is defined by the number of neighbourhoods between the other agent and the target.

PofAgent = the period that other agent has been in Passive mode, and

P = the maximum period of staying Passive.

If the condition $d \leq P - PofAgent$ (1)

is true, then the other agent will not be in an Active state by the time that the Infant agent reaches the vicinity of the target. The reason is that the distance between the target and the

other agent is less than the time that it takes for the agent to become Active. Therefore, by the time the other agent reaches the target, it is still in Passive state and cannot achieve the target. By using this formula, Infer agents can reason and predict when an agent will become Active. Founded on these predictions, Infer agents plan for the next step. The Infer agents' rule to move is different in Passive state than when it is in an Active state.

If there is no target in the field of movement and there is more than one target in its field of view, then a Passive Infer agent moves towards an empty cell that is in the vicinity of the target with the minimum of total number of agents in Active states. Infer agents consider other agent's mental state as Passive if and only if all of the following conditions are met:

1) If Infer agent has observed the competitor agent achieved a target.

2) The other agent's states does not change to Active by the time it reaches the target.

$$d \leq P - PofAgent \qquad (2)$$

3) The other agent's period is less than or equal to the Infant agent's period. Therefore, Infer agent becomes Active before the other agent.

$$PofAgent \leq PofInfer \qquad (3)$$

In the case where there is no target in Infer agent's field of view, then the Infer agent moves to an empty cell randomly. Thus, in situations where there is more than one target in the field of view and no target in the field of movement, the general rule is that Infer agents need to search for the target with the minimum competitors. This criterion is necessary but is not sufficient. The distance between the target and the Infer agent needs to be less than or equal to the difference between the general period and the other agent's period. Moreover, once Infer agent is Passive it is necessary to consider which agent will first become Active in the next step. Therefore, it is important to distinguish that the Infant agent's period needs to be less than the other agents' period.

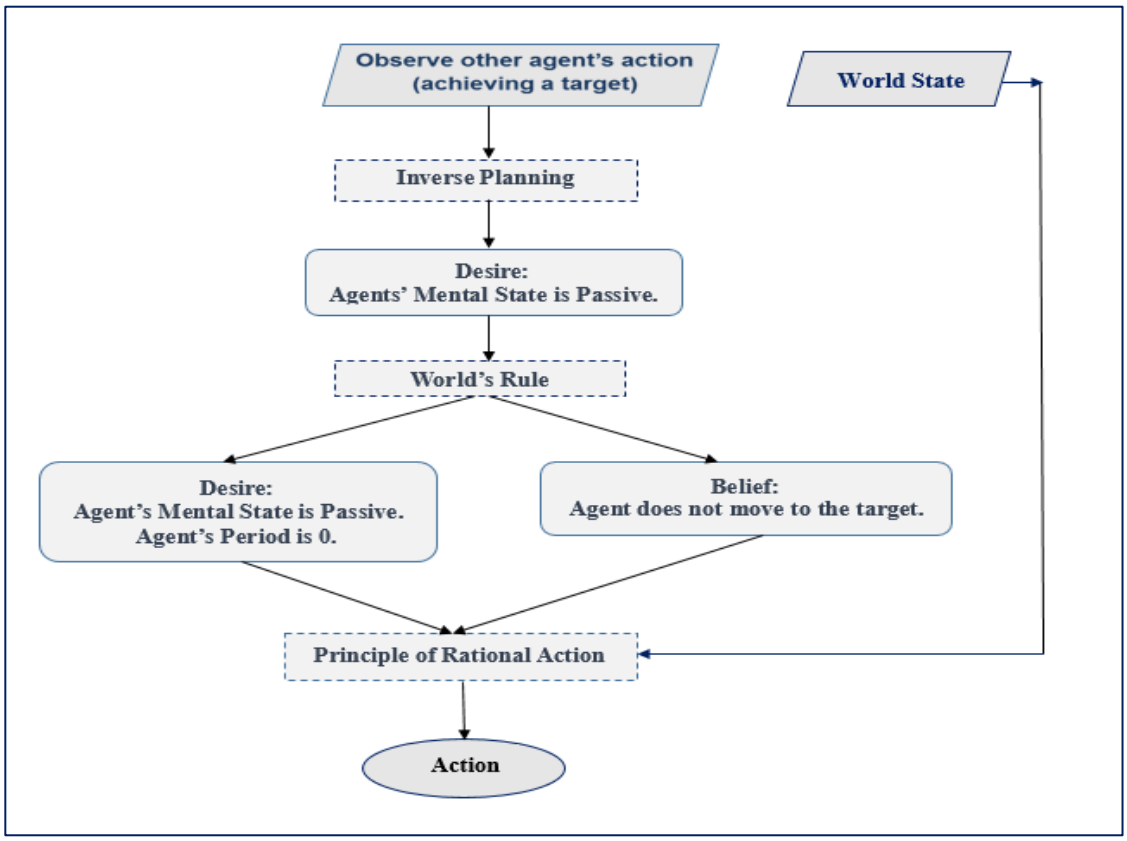


Figure 33. Infer Agents' inferences flowchart

3.2.4.6 Reality Agents' (Reality) Strategy

The Reality agents' strategy is similar to that of the Infer agents. The only difference is that the Reality agents have directly access to other agents' real mental states and their period of staying Passive. Therefore, the Reality agents do not infer the mental state of others. They do not require to store others' mental states information. In essence, their strategy and the ability to understand, analyse, plan and act based on others' mental states is similar to Infer agents without using the infer system. Reality agents' design and analysis is beneficial to evaluate the infer system efficiency and its role in agents' performances.

3.2.5 MSM Implementation

The MSM design comprises two abstract classes; Society and Agent. Agents are moving to achieve their targets in a grid space where Society contains all of the other classes

including Agent, Target, and Grid. Figure 34 shows the MSM class diagram including the classes, the relationships between them and their methods. The composition association between Society class and other classes, the black diamond sign in Figure 34, shows that Society class is a container for objects of other classes. Agent class is an abstract class including three main subclasses: MinToM Agent, Infer Agent and Control Agent. The 1-1 association between Agent class and Grid class demonstrates that every agent is placed in one and only one cell of the grid. Likewise, the 1-1 association between Target class and Grid class indicates that each target can exist in one cell.

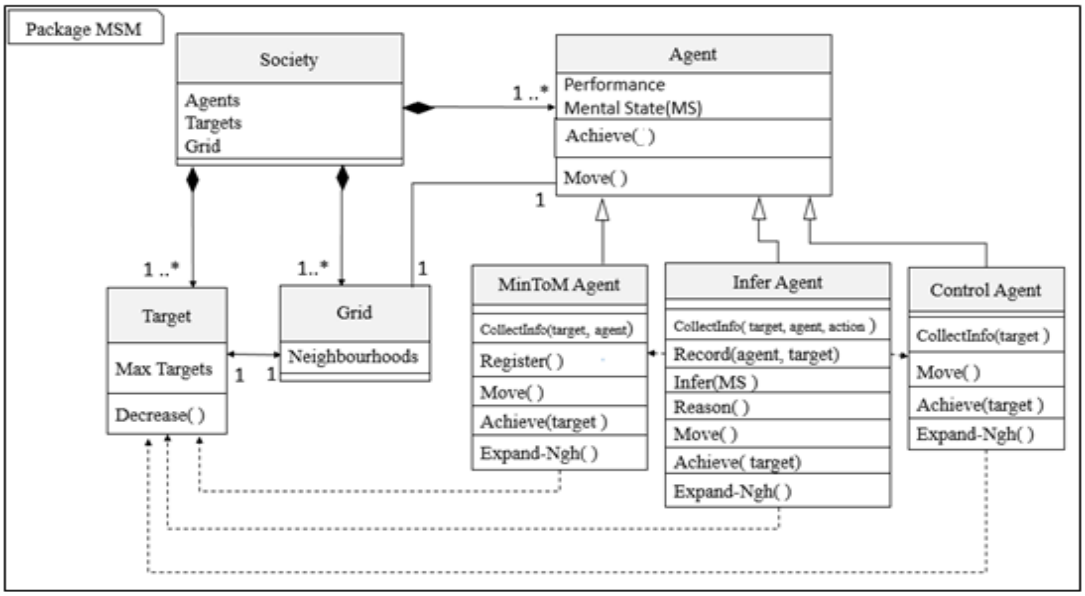


Figure 34. Class Diagram of MSM

The dependency relationship, shown by dots in Figure 34, illustrates the flow of the information, about the target, from MinToM Agent to Grid class as they achieve a target. A similar dependency relationship occurs from the Infer Agent class to the Grid class and also from the Control Agent class to the Grid class. The subclasses of the Agent class inherit its methods. In addition, they have their own specific methods. For example, Infer Agent class methods includes CollectInfo, Record, Infer, Expand-Ngh and Reason whereas MinToM Agent class methods are the CollectInfo, Register and Expand-Ngh methods.

Notice that although the name of some of the methods are the same and they are supposed to perform the same function, yet, they work differently with different parameters. The Decrease method of the Target class removes the target which has been achieved by an agent from the grid. Figure 35 shows a screenshot of the MSM simulation run.

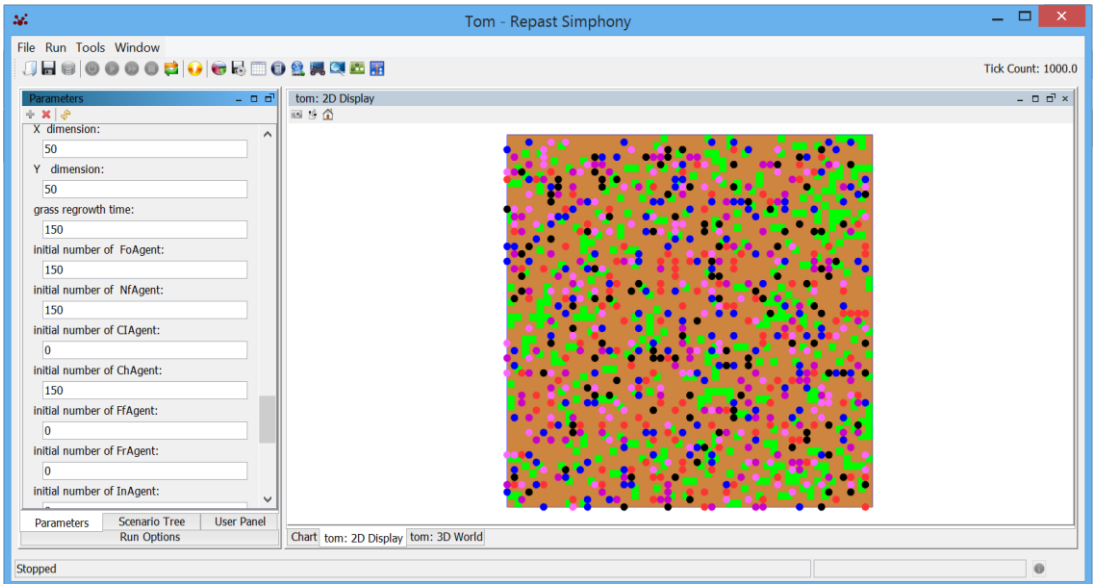


Figure 35. Screenshot of MSM run

3.2.6 MSM hypotheses and predictions

The three main agents in MSM are competent of different levels of theory of mind ability. Agents which are capable of inferring others' mental states (simple theory of mind ability), agents which are able of tracking others' beliefs (minimal theory of mind ability) and agents with no theory of mind ability. Thus, the level of theory of mind is an independent variable and agents' performance is a dependent variable in the simulations.

The hypothesis is that agents' level of theory of mind has an impact on agents' performances in achieving their targets in a competitive society. The ability of understanding others' mental states is an important factor in agents' performances. Therefore, the expectation is that agents' performances correlate with their level of theory of mind ability. The most efficient agents are those with theory of mind ability. Next are

the agents with minimal theory of mind ability and the agents with no theory of mind ability are least effective.

The general analysis of MSM is based on agents' performances in achieving their goals. Simulations have four parameters with a diverse range of values including extremes. The variety of scenarios result in a significant amount of data with different scales requiring normalisation. Thus, a normalised formula is introduced to organise the data in a standard scale which enables the comparison between various situations in the simulation. The next step of analysis is concerned with the largest difference in performances between agents. For this purpose, the minimum and maximum baseline of agents' performances is used to scale the differences between [0, 1]. In sum, the MSM analysis is divided into these three categories; the general performances, the normalised performances and the performance differences based on the minimum and maximum baselines. Each of this categories is explained in four subcategories based on each of the four parameters; number of agents, number of targets, period and neighbourhood.

The simulations have been conducted with a variety of parameter values. These parameters have been systematically changed and the simulation runs demonstrate clear results and interpretations. Therefore, using statistical analyses is not necessary for the aim of this simulation.

3.3 The MSM SIMULATION RESULTS

The results of MSM simulation have created a large amount of data for evaluation and analysis. In this respect, some classification is made to organise the results in a clear way. First, the simulation run consists of two initialization setups:

- Setup with all types of agents in one environment that is called 'Altogether run'.
- Setup with each type of agent in a separate environment with the same values of parameters that is called 'Single (Alone) run'. Thus, there are two results sections:

Altogether run and Single run. This section starts with the simulation graphs of both setups to check their results and analyse their differences. Then, the section continues only with the Single run simulations.

Second, each of these sections includes five different parameters and their corresponding graphs. In addition, running the MSM simulation with a variety of parameter settings provides more reliable patterns and diverse results regarding the agents' performances.

The parameters and their randomly selected values are:

- Number of targets (T) = 50, 100, 150, 200, 250, 300,400, 800, 1200
- Number of agents (N) = 50, 300, 400, 500, 600, 700, 800, 1200, 1600, 2000
- Number of periods of staying Passive (P) = 1 to 10, 12, 15, 17, 21
- Field of view (Ngh) = 2, 3, 4, 5, 6
- Time Step = 1000 Ticks

The average values for each set of parameters after running the simulation 10 times determine their values for the results. One practical approach to evaluate the performance of agents is to calculate the total number of time steps that Active agents fail to achieve a target. Such calculations for each type of agents in a similar environment indicates the performance of agents.

The analysis of the results starts with the 'general' graphs of Altogether and Single simulation runs to provide a general insight of the agents' performances. It is followed by examination of the effect of parameters of field of view (Ngh), number of agents (N), periods of staying Passive (P) and the effect of number of targets (T), in that order. It continues to analyse the normalised data based on the parameters N, T, P and Ngh. Then there is an evaluation of the main differences between agents' performances. In addition, the differences of Infer and MinToM agents based on Reality and Control agents' differences are explored more deeply. The results of Infer agents with a modified action,

and the general verification and validation of the model are explained at the end of this section.

3.3.1 MSM Agents' Performances

One main objective of MSM is to assess the relationship between the characteristics of agents' levels of theory of mind and their efficiency in achieving their targets. The agents' theory of mind capabilities have been introduced through their strategies. Thus, computing agents' performances, the link between their capabilities and their success and analysing the results in respect to their own and others' mental states is the aim of this section. In general, there are two possible situations for Active agents, as demonstrated in Figure 36. In the first situation, Active agents achieve their targets. In the second situation, agents' desire is to achieve the target but they fail.

Figure 36. Two possible situations for Active agents

The total number of time steps in which Active agents fail to achieve a target is the performance measurement in the simulation (X). Therefore, as the value of this measurement increases, the efficiency of agents decreases.

The significant amount of data produced from the simulation run demonstrates the agents' performances with a variety of different parameters, which demonstrates the reliability of the results. The general pattern of agents' performance efficiency in ascending order is: Random agents, Food agents, Control agents, MinToM agents, Infer agents and finally Reality agents. The majority of results evidence this throughout the simulation run. An example of this overall pattern is the Altogether simulation run with N=350, Ngh=2, P=4, T=250, shown in Figure 37 and Figure 38. The differences between Random agents and

other agents is large, thus Random agents were removed from the graph in Figure 38 to assess other agents' performances more closely. The graphs show the performance of Random agents is the worst and Reality agents demonstrate the best performance. Infer agents perform approximately as high as the Reality agents. This suggests that the Infer agents' strategy is a reliable method for inferring others' mental states in this world.

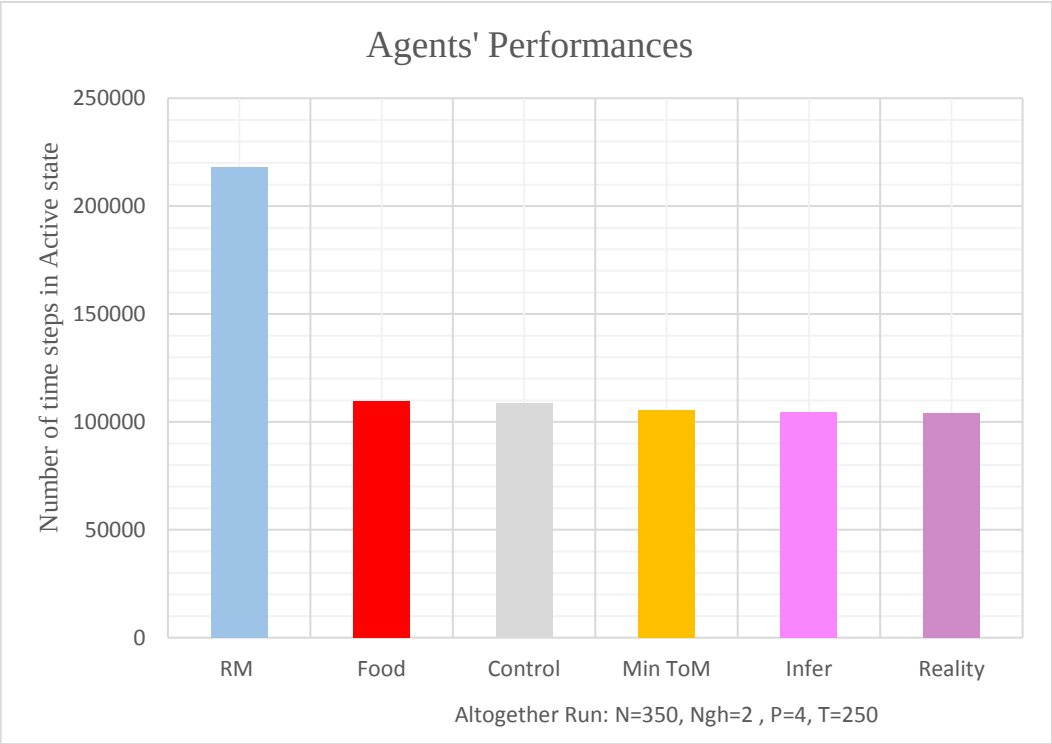


Figure 37. The results of Altogether simulation
Set up of 350 agents, 250 targets, Ngh = 2 and P = 4, in 1000 Tick.

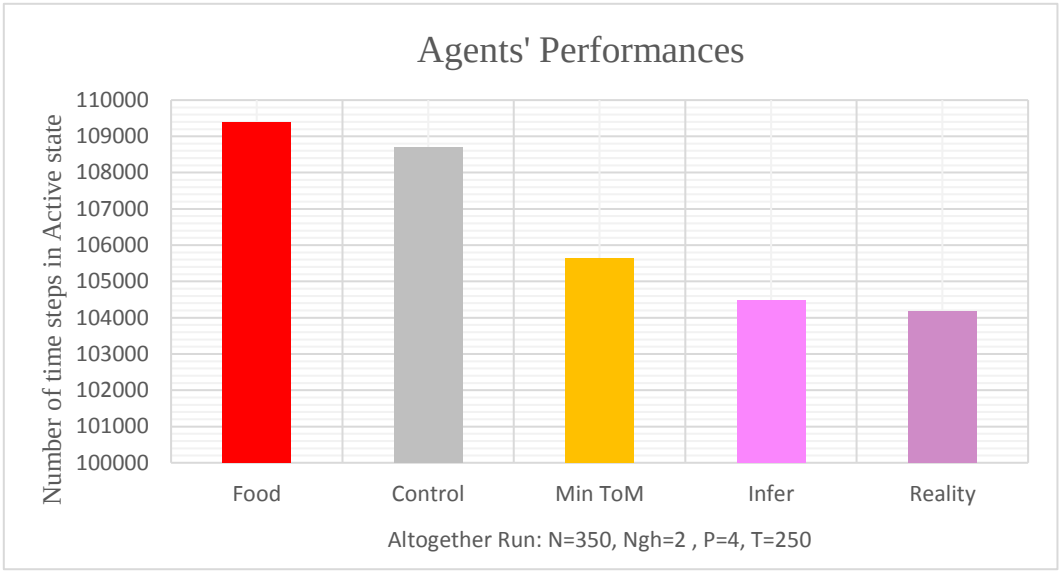


Figure 38. The results of Altogether simulation without Random Agents
Set up of 350 agents, 250 targets, Ngh = 2 and P = 4, in 1000 Tick.

The analysis of parameters in the environment is crucial to understand the results of agents' performances; the total number of agents in the environment is 1750, which is high, and there is a low number of targets, which is 250, in 2500 cells. The ratio between the numbers of targets to agents (1/7) indicates that on average, for every seven agents there is only one target available in the environment. The agents' field of view is 2 which is equal to observing 24 cells in its neighbourhood. In addition, the parameter value of the number of time steps staying Passive (P) has an indirect effect on this ratio due to the fact that each agent required to stay Passive for 4 time steps before becomes Active (P=4). Another example of agents' performances in the Single simulation context of MSM is shown in Figure 39 with parameters of N=1500, T= 250, P=4 and Ngh=2. The general patterns in these graphs are consistent with the previous graph's results. The pattern of the best performance of agents in descending order is: Reality, Infer, MinToM, Food, Control and Random agents. Control agents' efficiency is slightly less than that of the Food agents as

the Control agents are unable to expand their field of view (Ngh=2) to a wider area in this setup.

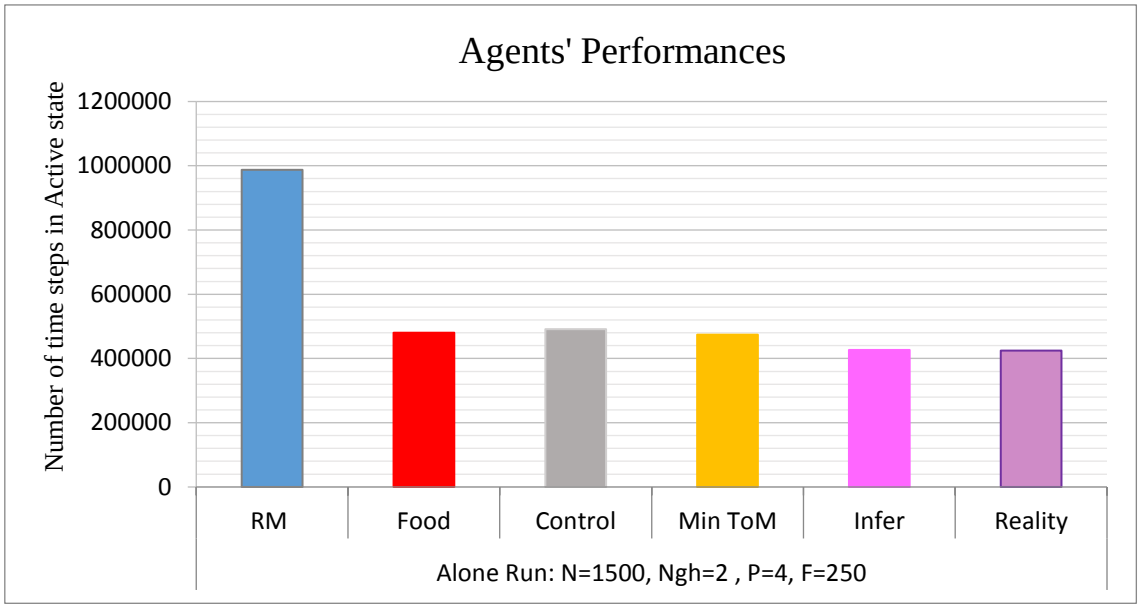


Figure 39. Agents' Performances in Single simulation run
Set up of Ngh=2, N=1500, P=4, T=250 in 1000 Ticks (10 times run).

In the following sections, the analysis of the agents' performance continues in greater detail and a more organised style based on different parameters' effect on agents' performance.

3.3.1.1 The effect of Field of View (Ngh)

Figure 40 and Figure 41 illustrate the effect of Ngh on agents' performances. This effect does not apply to the Random and Food agents because field of view has no role in their strategies and actions. Thus, Random and Food agents are not included in these graphs.

By increasing the field of view, the performance of the Control, Min ToM, Infer and Reality agents consistently increases. However, there are some limitations to this; for example, the Control agent shows a sharp increase from Ngh=2 to Ngh=3 but not the same applies from Ngh=3 to Ngh=4. Due to the distribution of food within the environment, it is highly unlikely that an agent will reach a target at Ngh=4 because another agent will have consumed it. In other words, for more than Ngh=3, the target is no longer available. By

increasing the field of view, the Infer and Reality agents have a systematic increase in their performances up to Ngh=4.

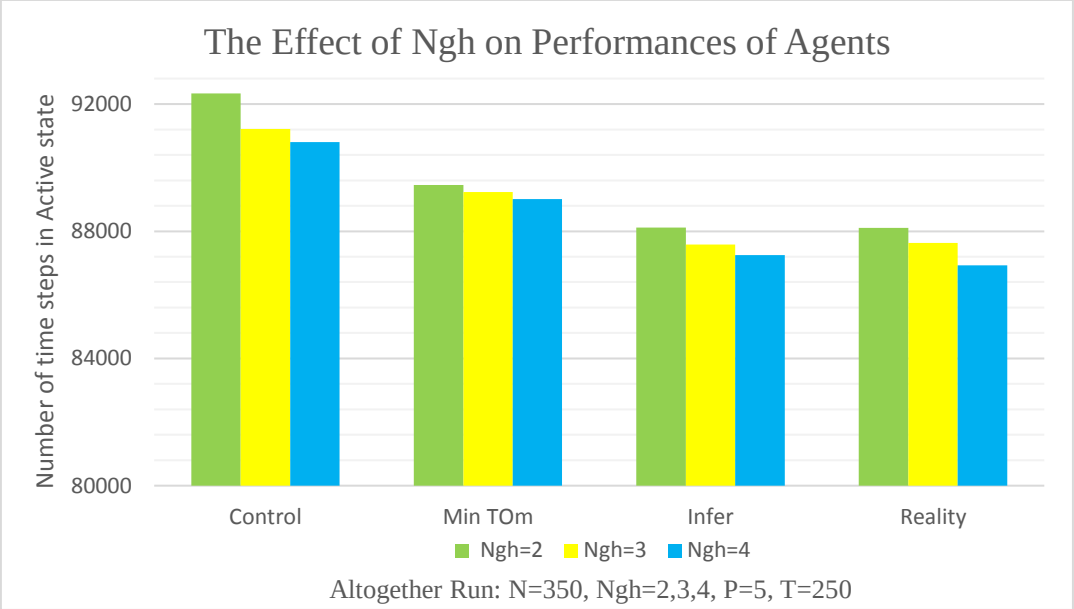


Figure 40. The results of Altogether simulation based on Ngh
Set up of N=350 agents, T=250 targets, Ngh= 2, 3, 4 and P= 5, in 1000 Ticks.

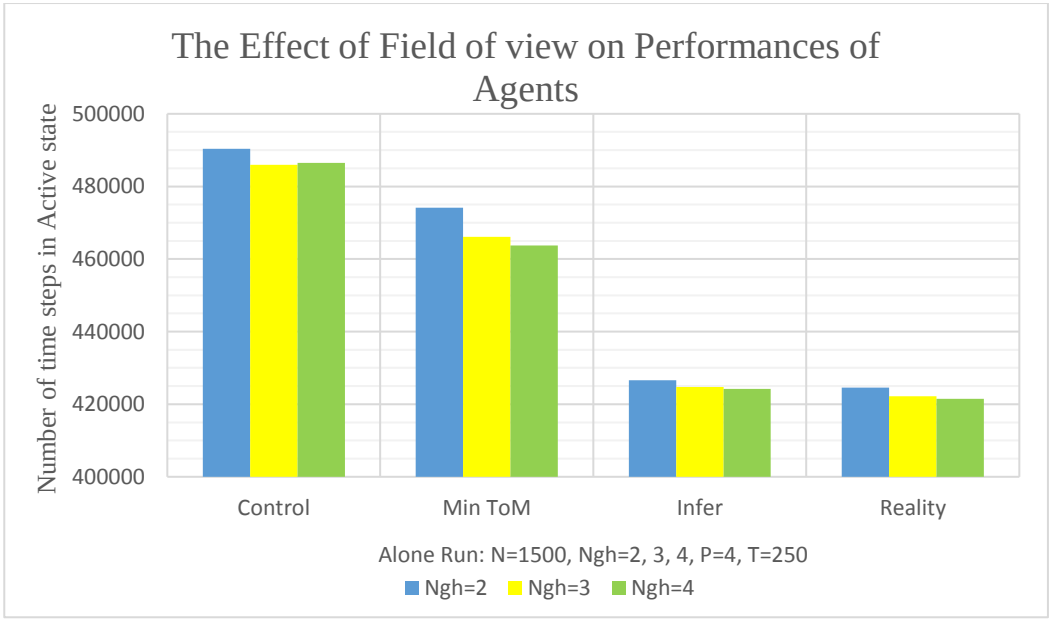


Figure 41. The results of Single simulation run based on Ngh
Setup of N=1500 agents, T=250 targets, Ngh= 2, 3, 4 and P= 4, in 1000 Ticks.

3.3.1.2 The effect of Number of Agents (N)

The effect of the number of agents on agents' performances with setups of: N=250, N=300, N=350, N=400 and Ngh=3, P=5, T=250 of Altogether simulation run initialization is shown in Figure 42. This figure shows that as the number of agents increases, their performances consistently decreases which demonstrates that as the population increases the possibility of achieving a target with the same resources decreases.

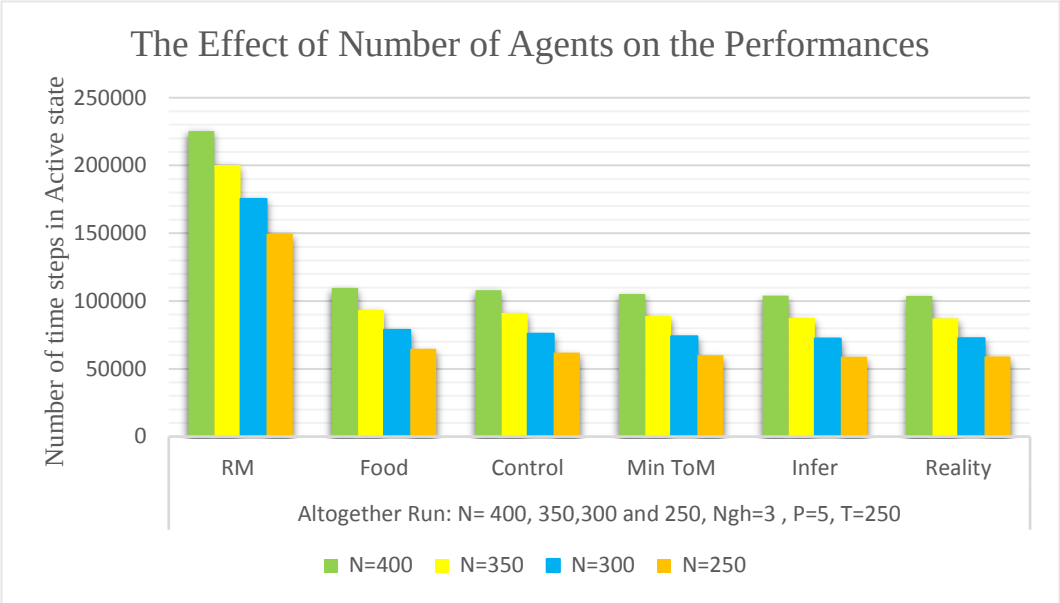


Figure 42. The results of Altogether simulation run based on N

Similarly, the effect of the numbers of agents with: N= 1000, 1250, 1500, 1750 and 2000, Ngh=3, T=250, P=5 in a Single simulation environment is shown in Figure 43. It indicates that as the number of agents in the environment increases, the total number of time steps that each type of Active agent fails to achieve a target increases. This result includes Random agents, and intriguingly indicates that Random agents' performances decreases as the number of agents increases, regardless of their lack of strategy. However, other agents' performances are affected less significantly than Random agents.

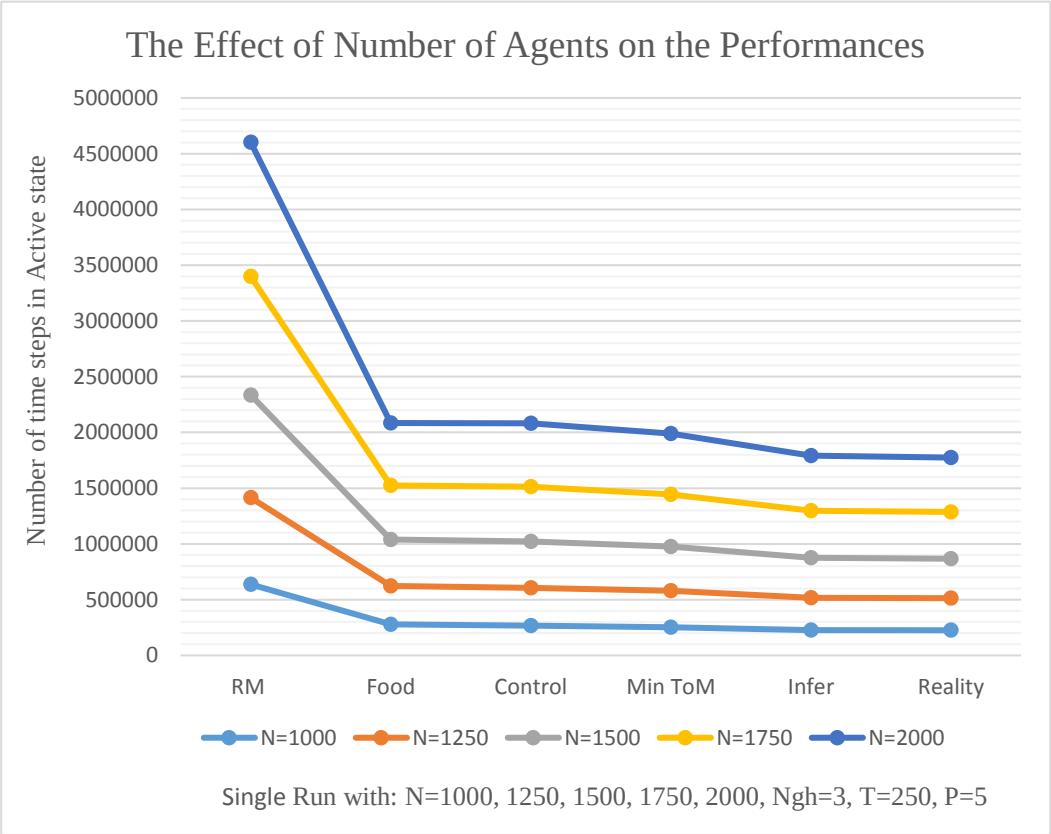


Figure 43. The results of Single simulation run based on N

Noticeably, the agents' performance for N=1000 is almost twice as efficient as for N=2000 in a similar situation of targets and parameters. In other words, as the number of agents doubles, the efficiency of agents halves.

3.3.1.3 The effect of Period of staying Passive (P)

The graphs in this section concern the parameter P, the period of staying Passive, and its effects on the results of the simulation by varying P from 1 to 10. As expected, the results confirm that the agents' performances improve as P increases. These results are comparable to the results of previous graphs; demonstrating that the efficiency of agents' performances, in descending order, is: Reality, Infer, Min ToM, Control, Food and Random agents.

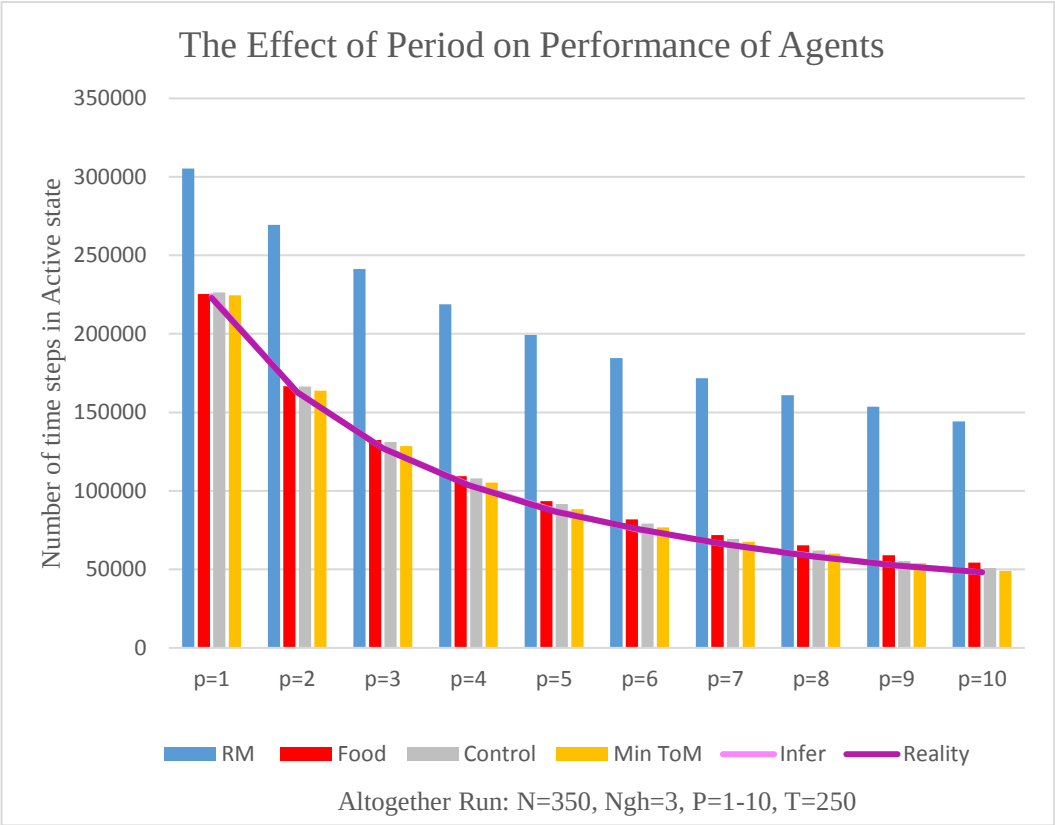


Figure 44. The results of Altogether simulation run based on P

Figure 44 shows that as the period increases, the agents' performances improve because their mental states stay in Passive state for longer. However, the percentage improvement varies between different types of agents. For example, Reality, Infer, and MinToM agents' performances improve by around 4.8%, 4.6%, and 4.5%, respectively, while Control, Food and Random agents' performances improve by 4.5%, 4.1% and 2.6%, respectively. Since the number of agents are as low as 350, this percentage improvement varies with the number of agents.

Similarly, the effect of altering the parameter period for the Single run with N=2000, Ngh=3, T=250 and P=1 to P=10, is shown in Figure 45. Similarly to the Altogether run results, as period increases, agents' performances increase due to staying in the Passive state longer. In addition, it demonstrates that the order of the agents' performance in terms of efficiency is consistent with the earlier graphs.

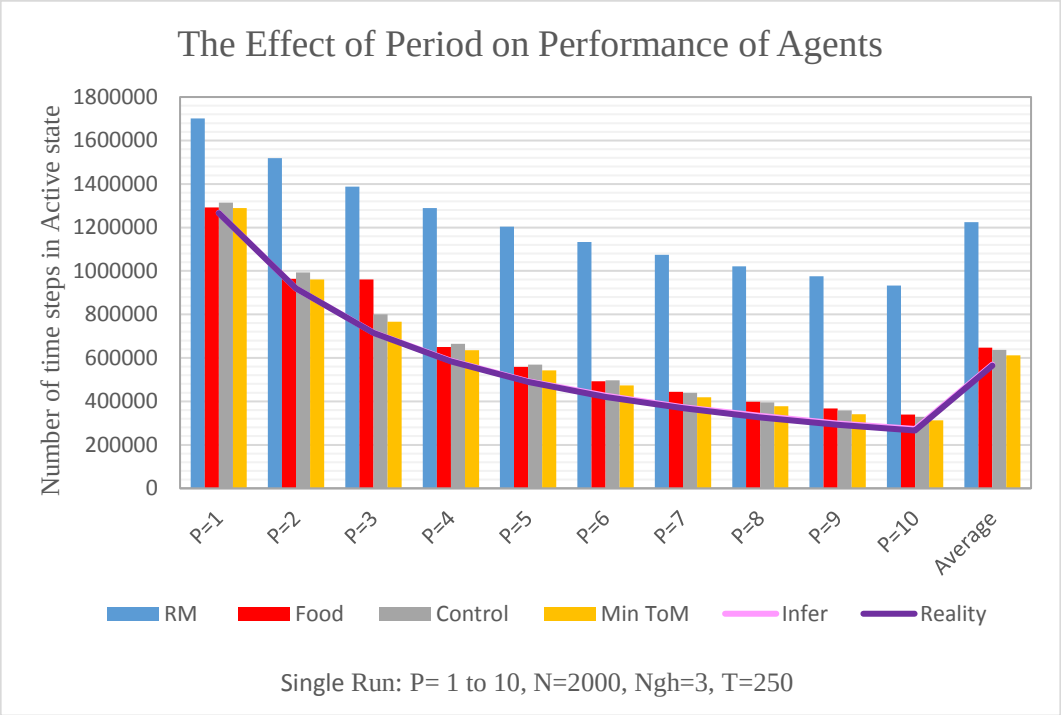


Figure 45. The results of Single simulation run based on P

3.3.1.4 The effect of Number of Targets (T)

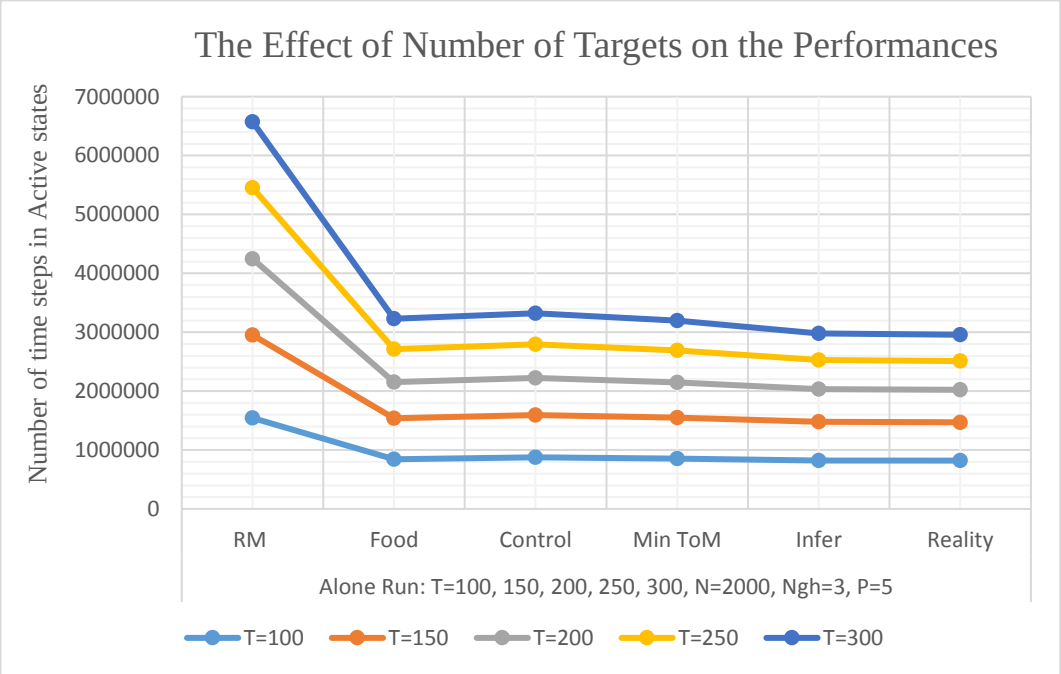


Figure 46. The results of Single simulation run based on T

One important parameter of the simulation is the number of targets. The effect of the number of targets on the performances is shown in Figure 46, with the parameters of N=2000, Ngh=3, P=5, T=100 to 300 in a Single setup. The graph indicates that the Reality and Infer agents perform more efficiently than others, as expected. However, the Food agents' performance is higher than the Control agents and also higher than MinToM agent when the number of targets is low (T=100). The reason for this is that when the number of targets are approximately one in 250 cells, the level of uncertainty is high for MinToM agents, which makes them unable to use their strategy to track others' field of view. Thus, they are less efficient because their strategy criteria are unlikely to be fulfilled. The parameters in this graph demonstrate a low number of targets in the environment; the ratio between the number of targets and the number of agents is between 1/20 and 3/20, which is very low. All of the graphs to this point are only a snapshot of the MSM simulation run. The main results and the organised data are explained in the next section.

3.3.2 The MSM Normalised Results

Each of the MSM simulation runs produce a set of data. This significant amount of data can be shown in a large number of graphs to illustrate the patterns which, when analysed, answer the primary questions posed to MSM. For the purpose of comparison between various situations in MSM, it is necessary to normalise the data. Therefore, by operating normalisation, the data is organised in a standard scale which reflects a variety of situations and clarifies the relationships in the data.

3.3.2.1 Normalised Formula

The main value from the simulation run for each type of agent is the total number of time steps in which Active agents fail to achieve any target, and is called X. In fact, the value of X is a measurement for evaluating the agents' performances. In every simulation run, the parameters, N, T, P and Ngh, determine the minimum and maximum performances of

agents (the minimum and maximum X values). Table 5 shows the parameters and their values which have been systematically selected by considering extreme values for analysing the normalisation.

No Of Agents	No of Targets	Period	Ngh
50	50	1	2
400	400	2	3
800	800	7	4
1200	1200	12	6
1600		17	
2000		22	

Table 5. Parameters and the values for MSM simulation

By multiplying the number of different values each parameter can have, we can find that there are (6*4*6*4=) 576 different combinations of settings we can use to represent and analyse the results.

Although X does show different agents' performance efficiency within a simulation run, values of X from simulations with different parameters cannot be directly compared to each other. Therefore, to accomplish a standard evaluation of agents' performances in all of these 567 simulations, it is essential to use an approach to normalise the data. A new approach to normalise the value of X has been calculated. The worst scenario, which illustrates the agent's failure to achieve any target, happens when the value of X is at its maximum.

Consider "P" is the period of staying Passive and "Ticks" is the number of time steps. The ratio of (Ticks/P+1) determines the maximum number of time steps that an agent is Active.

Thus: $M = (Ticks/P+1) * N$ (where N is the number of agents)

Here, M indicates the maximum number of time steps that agents can be Active.

For example, if N=50, ticks=1000 and P=1, then M = (1000/2)*50 = 25000, which indicates these 50 agents spend a maximum of 25000 time steps in the Active state in a 1000 tick simulation.

Thus, the percentage of failing to achieve any target formula is:

$$\text{NormalisedX} = ((X-M)/M)*100$$

For example, consider two different cases of T=50 and T=800 with the parameters of N=50, Ticks=1000 and P= 1:

Case I : T=50

As X=49,887 (this number comes from the simulation results),

$$\text{NormalisedX} = [(49{,}877 - 25000)/25000]*100$$

$$\text{NormalisedX} = 99.50\%$$

The value of NormalisedX indicates that in 99.50% of time steps agents are in Active states. This indicates that agents do not achieve many targets and their performance is low. The performance of agents is inversely proportional to the value of NormalisedX (i.e. as NormalisedX increases, the agents' performance decreases).

Case II: T=800

As X=25,111,

$$\text{NormalisedX} = [(25{,}111 - 25000)/25000]*100$$

$$\text{NormalisedX} = 0.44\% \quad \text{(to 2 decimals)}$$

In this situation, the probability that agents fail to achieve a target is only 0.44%. In other words, agents are 99.56% successful in achieving their targets.

NormalisedX values adjust X values to a percentage scale. Thus, NormalisedX values define the percentage of time that agents cannot perform effectively. Therefore, the higher

this percentage is, the lower the performances of agents. For example, by a comparison between the above cases, X does not simply show the discrepancy of the performances in the two cases. Whereas, by calculating NormalisedX, it is possible to compare agents' performances precisely. Undoubtedly, agents' performance is significantly lower in case I than in case II. This shows that by multiplying the number of targets by 16 (from 50 to 800), the agents' performance increases by over 99%.

3.3.2.2 Normalised Performance differences

Analysis of differences between agents' performances is a valid approach to assess their efficiency in more detail and identify the parameters involved in their improvement. The graphs of agents' performance differences with a variety of parameters' values show that there is a general pattern: the largest performance differences, in descending order, happen between MinToM agents and Infer agents, Control agents and MinToM agents, and finally Infer agents and Reality agents. Figure 47 is an example graph that shows the general pattern of these differences.

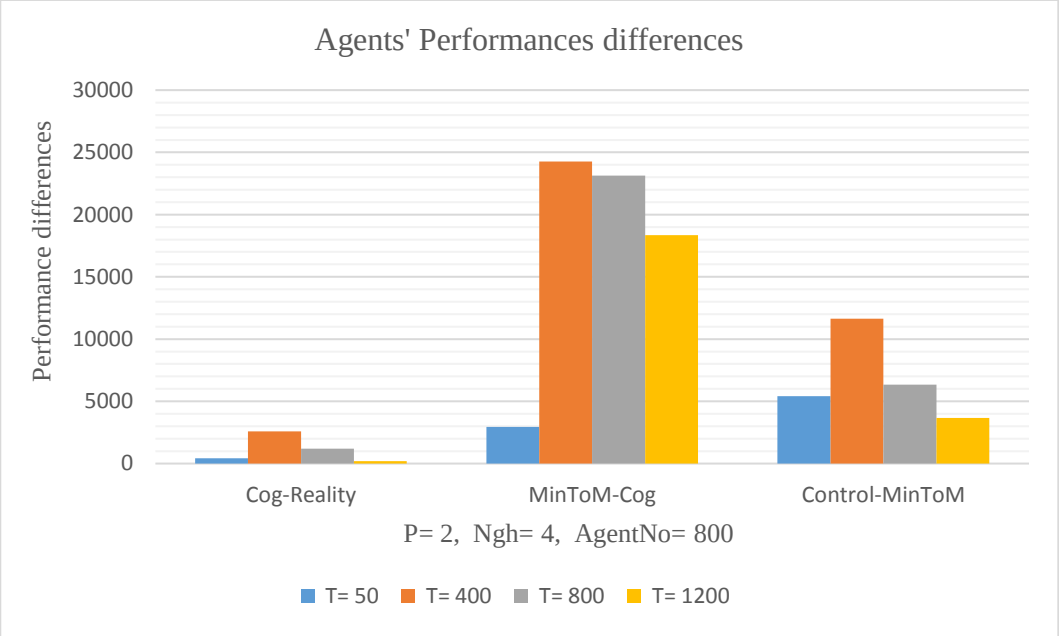


Figure 47. Agents' Normalised performance differences

The largest differences in agents' performance occurs between MinToM and Infer agents, excluding the Random agents. This indicates that the Infer agents' abilities to understand others' mental states are the critical point behind these results. Therefore, this section concentrates on MinToM and Infer agents' performance differences. Some of the conducted simulations graphs are presented in the following sections, and organised based on the effect of N, T and P parameters.

3.3.2.3 The effect of Number of Agents in Normalised data

The effect of the parameter N is shown in Figure 48 to Figure 51. All of these graphs depict normalised performances of MinToM and Infer agents based on the number of agents. Figure 48, with the setup of P=7, T= 50, Ngh=3 and N=50-2000, shows the performance differences between MinToM and Infer agents are very low. The reason is that the number of targets is 50, meaning the ratio of targets to cells in the world is 1/50, which is extremely low in relation to the size of the world (2500 cells). This extreme situation increases the uncertainty of the world, which consequently prevents the criteria that enable agents to practice their abilities and agents' strategies become inactive.

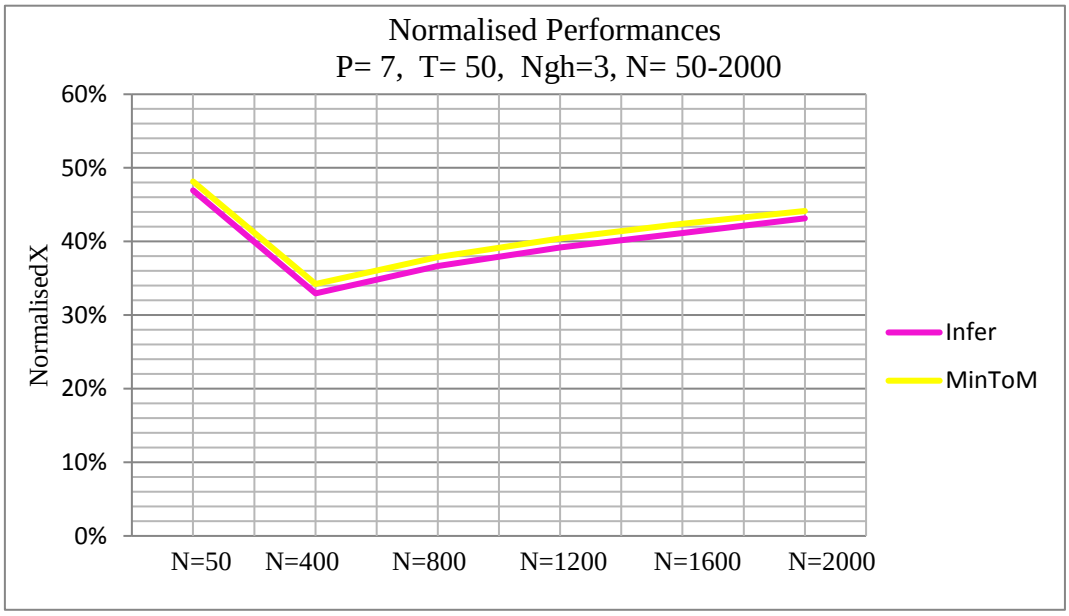


Figure 48. Normalised differences (Infer agents, MinToM agents) 1

In contrast, all of Figure 49 to Figure 51 show that the performance differences between MinToM and Infer agents in variety of situations dramatically increases. For example, Figure 50 shows a difference of nearly 2% with parameters of P=7, T= 800, Ngh=3, N=50-1200. Another example, Figure 49 shows an increase of approximately 3% for Infer agents performances when P=7, T= 400, Ngh=3, N=1600. However, Figure 51 illustrates that as the number of targets increases to 1200 and the agents' population increases from 50 to 800, the differences decrease because agents are able to achieve targets with less need for their specific abilities.

In conclusion, as the number of agents increases, the normalised performance differences of agents in almost all situations increases, except in extreme circumstances which cause unpredictability and uncertainty in the environment. In uncertain situations, when the number of targets is very low, and also in situations where the number of targets is higher than the number of agents, the criteria that enables agents to use their individual rules decreases.

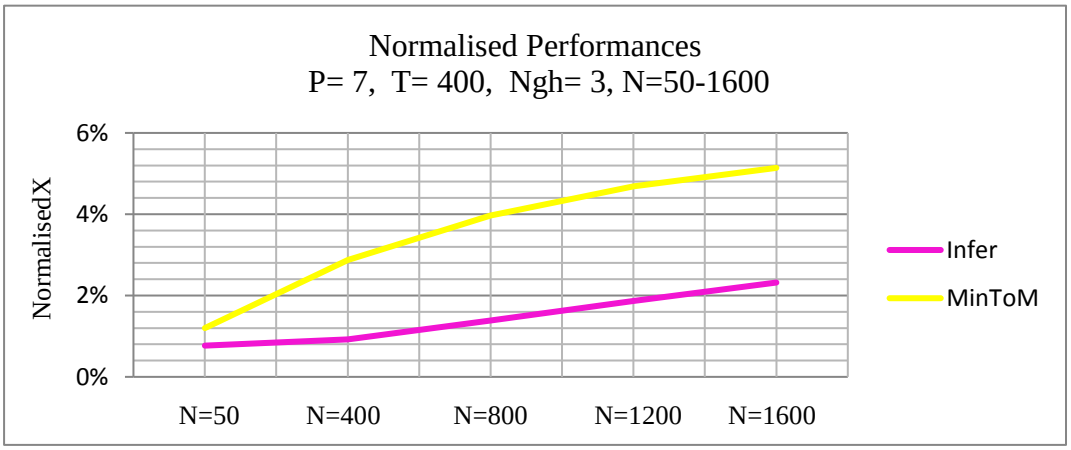


Figure 49. Normalised differences (Infer agents, MinToM agents) 2

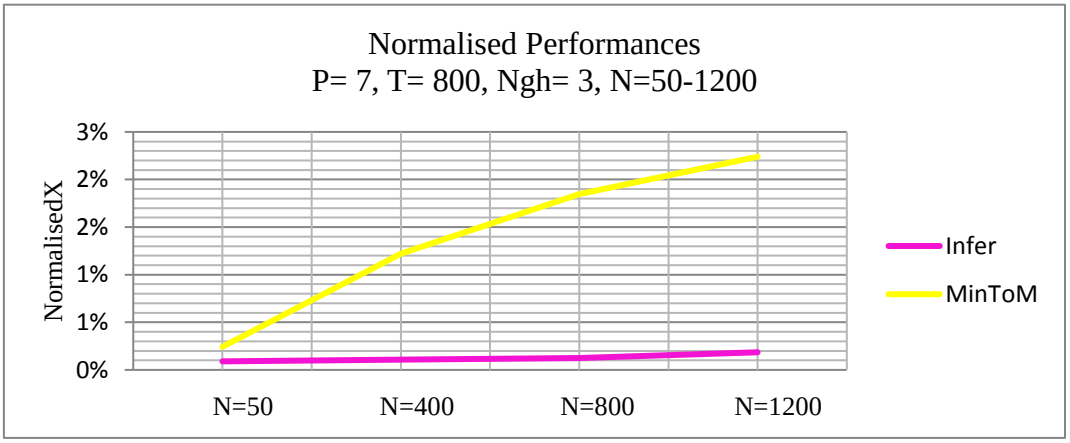


Figure 50. Normalised differences (Infer agents, MinToM agents) 3

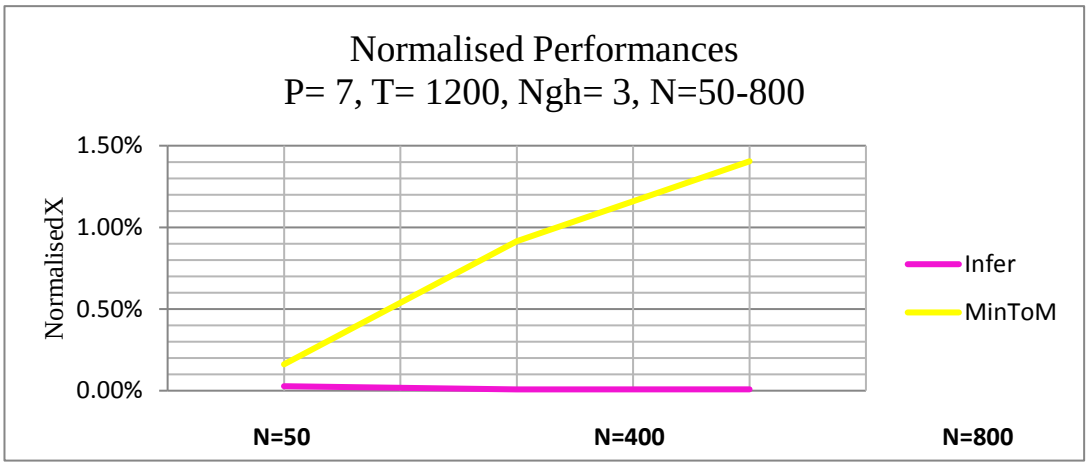


Figure 51. Normalised differences (Infer agents, MinToM agents) 4

3.3.2.4 The effect of Number of targets in Normalised data

Figure 52 to Figure 54 show the normalised performances of MinToM and Infer agents based on the number of targets. All of these figures show that the MinToM performances are significantly lower than the Infer agents' performances in a variety of situations. For example, there is a difference of nearly 2% within the parameters of P=7, T= 400, Ngh=3, N=400 in Figure 52. In the same figure, as the number of targets increases to 800 and then to 1200, the differences gradually start to decrease because the number of targets are larger than the number of agents. In other words, when the ratio between the number of targets

and the number of agents is greater than 1 (T/N > 1) and the ratio between the number of targets and the total number of cells in the environments is more than 1/8, then agents start to achieve targets using less of their abilities corresponding to theory of mind.

Noticeably, T/N is an important factor in agents' performances. This ratio has a great impact on the number of times that agents are able to use their strategies regarding their theory of mind ability. There are two extreme situations relating to the ratio which create two different situations; in the first case, the ratio is very low, therefore the number of targets are very low in relation to the number of agents, and in the second case when the ratio is high, this means that the number of agents are lower than the number of targets. In the first, case agents' performances are low and conversely in the second case, agents' performances are high. However, in both cases, agents are mostly unable to use their theory of mind ability and strategies. Hence, their performances do not directly reflect their theory of mind abilities in these two extreme cases. Moreover, the T/N ratio is not the only factor that directly influences the performance of agents. Certainly, other factors such as P, Ngh and N and T are involved.

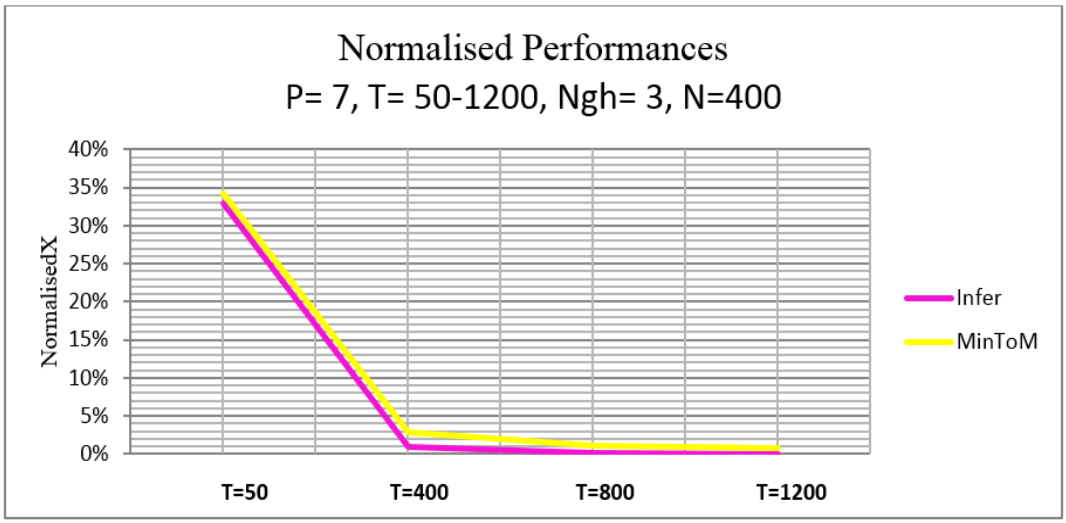


Figure 52. The first Normalised differences (Infer agents, MinToM agents) with T

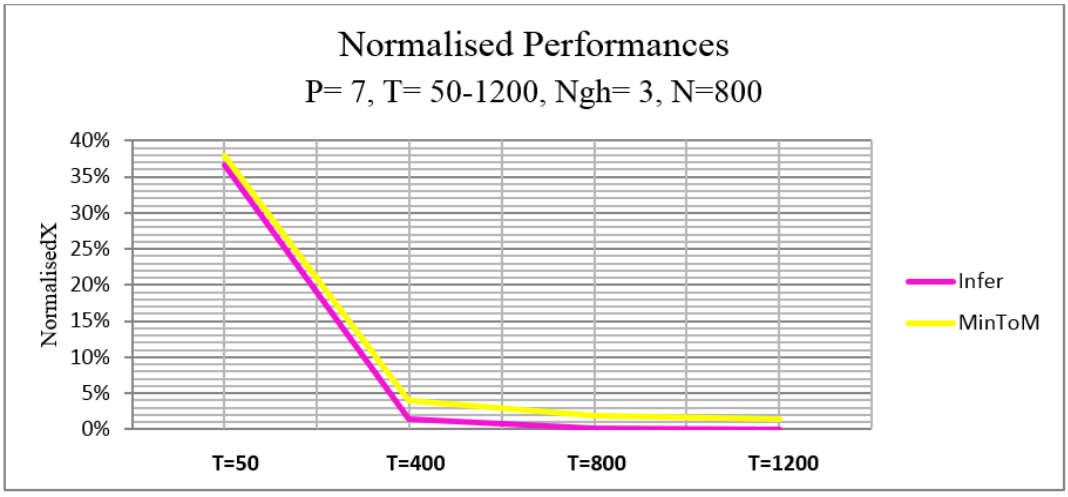


Figure 53. The second Normalised differences (Infer agents, MinToM agents) with T

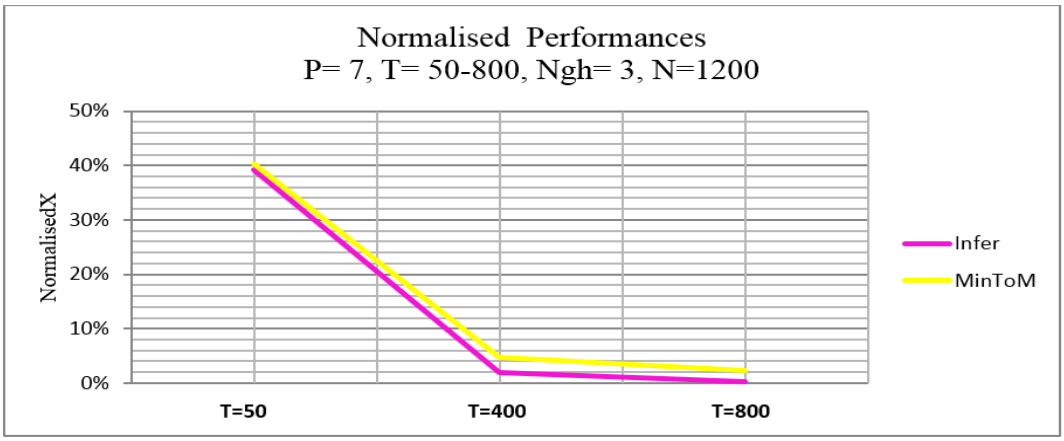


Figure 54. The third Normalised differences (Infer agents, MinToM agents) with T

3.3.2.5 The effect of Period of staying Passive in Normalised data

The graphs from Figure 55 to Figure 60 show the results of MinToM and Infer agents' performances with different parameter values including P=1, 2, 7, 12, 17 and 22. The main pattern in all of these figures shows that, generally, by increasing the parameter P, the normalised performance decreases. It also indicates that as P increases, the performance differences decreases. The explanation is that by increasing P, the time steps that agents stay in a Passive state increases and therefore agents achieve fewer targets.

All of the figures from Figure 55 to Figure 59 suggest that Infer agents' performances are more efficient than MinToM. Nevertheless, there are exceptions which occur in uncertain environments. For example, Figure 60 shows that both agents' performances are similar when the number of agents and the number of targets are equal to 50. Thus, it is less likely for agents to apply their rules and strategies in this situation.

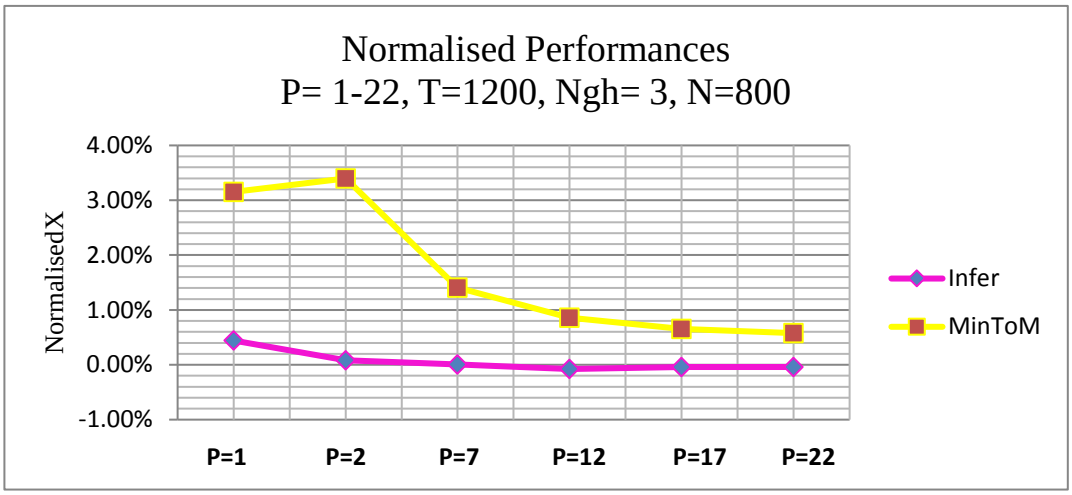


Figure 55. Normalised differences (Infer agents, MinToM agents) with P, (1)

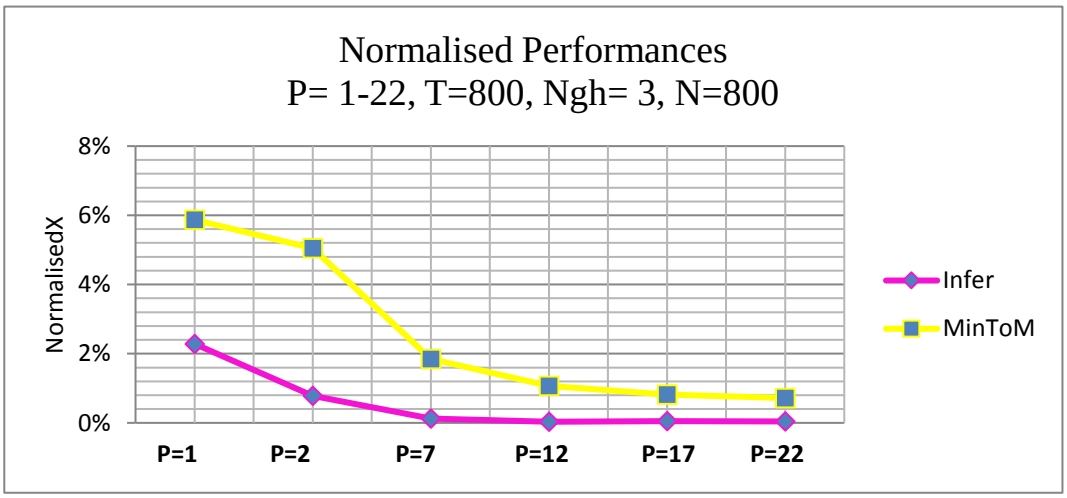


Figure 56. Normalised differences (Infer agents, MinToM agents) with P, (2)

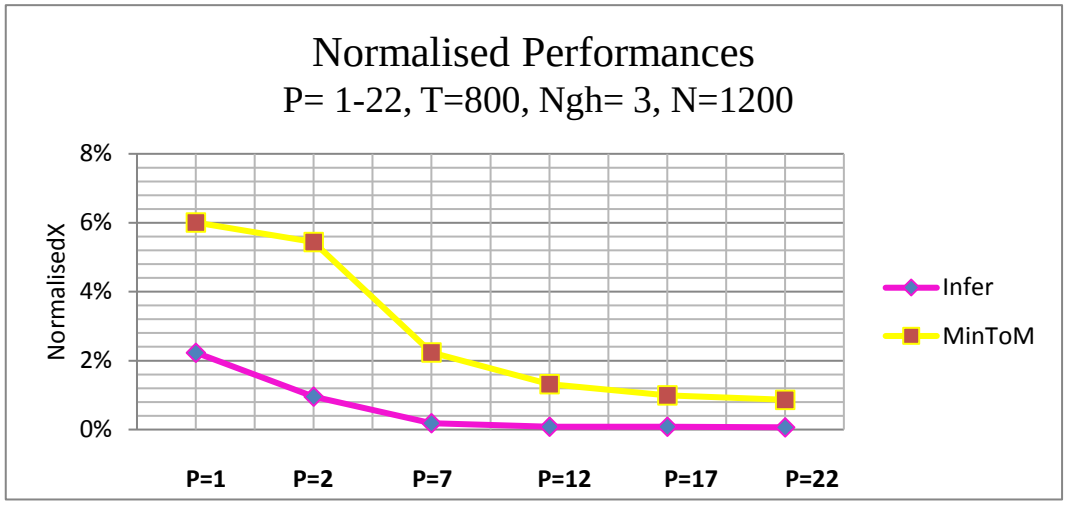


Figure 57. Normalised differences (Infer agents, MinToM agents) with P, (3)

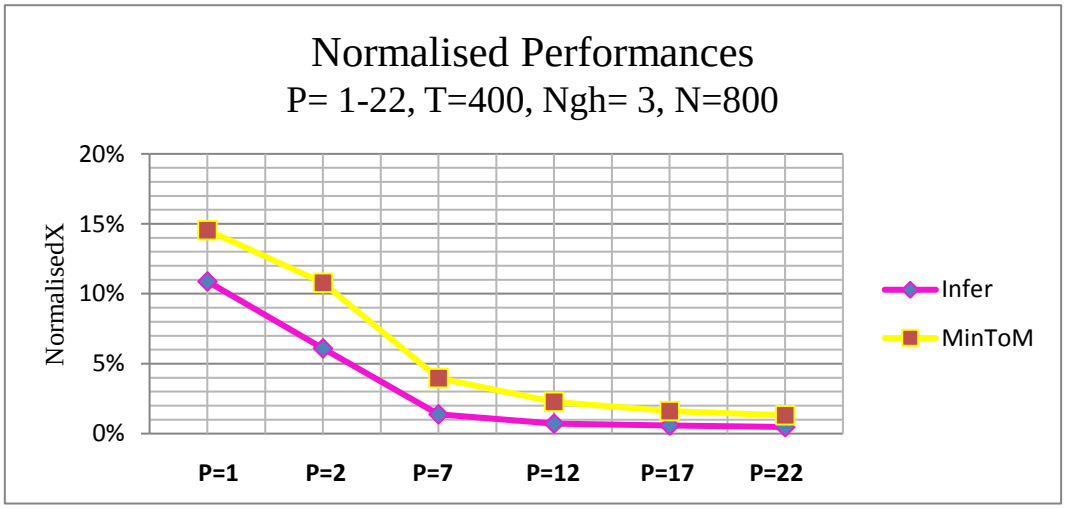


Figure 58. Normalised differences (Infer agents, MinToM agents) with P, (4)

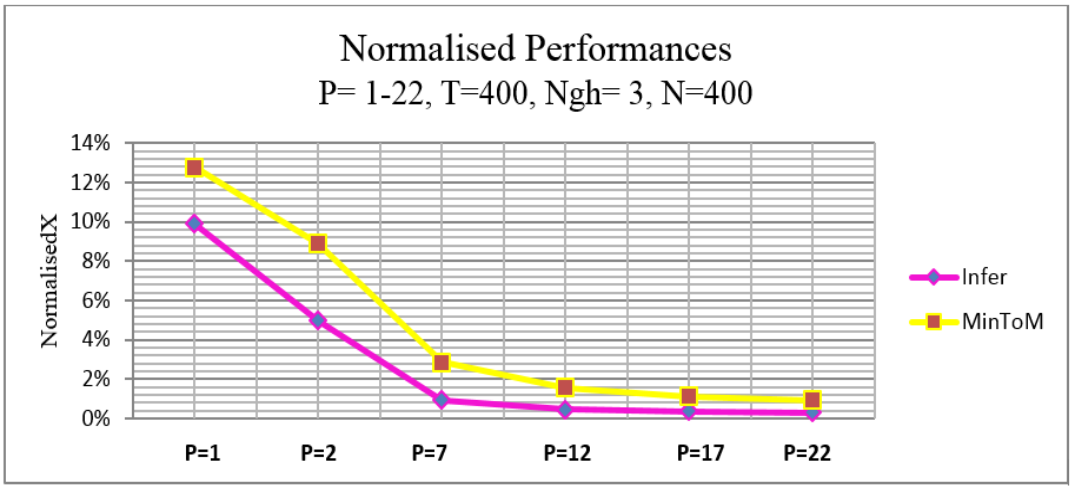


Figure 59. Normalised differences (Infer agents, MinToM agents) with P, (5)

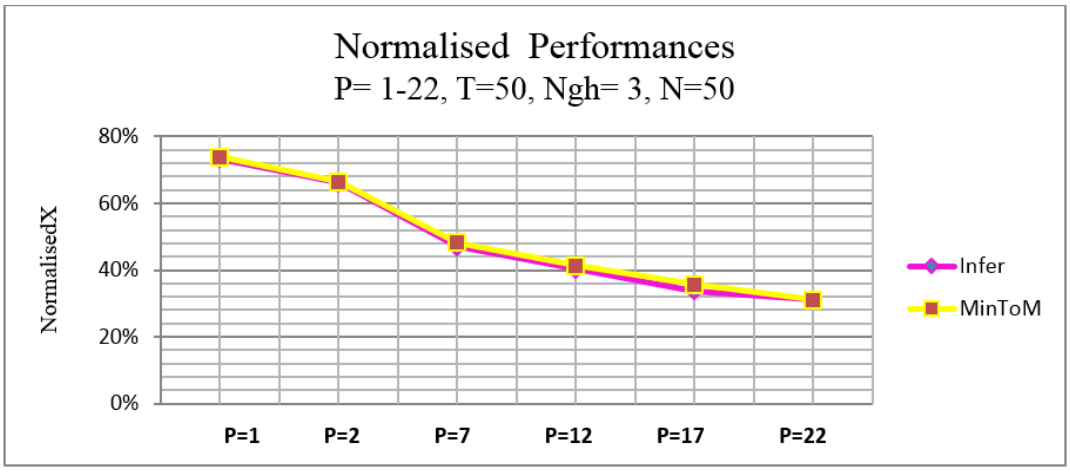


Figure 60. Normalised differences (Infer agents, MinToM agents) with P, (6)

3.3.2.6 The effect of Field of View in Normalised data

The effect of field of view on Control, MinToM, Infer and Reality agents' normalised performances is shown in graphs from Figure 61 to Figure 84. The effect of field of view depends on the number of targets in the environment. Thus, the results are divided into three categories based on the density of the number of targets in the environments; low-density, medium-density and high-density.

Low-density Environment

The most effective impact of field of view arises when the density of targets in the first neighbourhood is very low. In other words, when there are no targets in the first neighbourhood. Therefore, agents need to use their vision to search for a target at a further distance. Agents do not expand their field of view unless it is crucial. For example, in the case where there is a lack of targets in the nearest neighbourhood, agents expand their field of view to find a target at a further distance. Thus, this enables them to plan for the next time steps (by applying their strategies) and while there is no target in the first neighbourhood, they will move towards the nearest target in their field of view with the fewest competitor agents around it, which might enhance their performance.

In an environment with a low-density of targets and of agents, causing an uncertain situation, the role of Ngh parameter influences the agents' performances. Figure 61 shows that all types of agents' performances improve by between 64% and 70% as the field of view increases from 2 to 6, with parameters of P=7, T=50, Ngh=2, 3, 4, 6, N=50. Intriguingly, in this situation, agents' abilities in relation to others' perspectives are less applicable in the current time step, due to insufficient targets in the first neighbourhood, regardless, they are still able to follow the targets at the further distance by expanding their field of view. Furthermore, by increasing the number of agents to N=400, Figure 62 shows that agents' performances improve by approximately 6% as the field of view increases from 2 to 6, with the parameters P=7 and T=50. Similarly, for N=800, Figure 63 illustrates 4% improvement for MinToM agents and Infer agents' performances, whereas for Control agents, it is 3% as the field of view increases from 2 to 6. It is striking that from Ngh=2 to Ngh=3, the agents' performance changes more than from Ngh=4 to 6. This indicates that the expansion of field of view from Ngh=2 to Ngh=3 occurs because the likelihood of having a target in Ngh=3 is higher.

However, in a low-density environment, as the number of agents increases, the effect of field of view decreases. For example, a comparison between Figure 61 and Figure 63 indicates a decrease in the effect of field of view in agents' performances from 65% to 3% while the number of agents increases from 50 to 800 and the field of view increases from 2 to 6. Similarly, in a low-density environment, as the number of targets increases, the effect of field of view declines. For example, a comparison between Figure 61 and Figure 66 shows that Infer agents' performance improves by 70% and 0.5%, respectively, when the number of targets is increased from 50 to 400 and field of view is increased from 2 to 6.

In contrast, in a low-density environment, as the period increases, the effect of field of view increases. In general, the ability to expand agents' field of view will improve agents' performances in low-density environments.

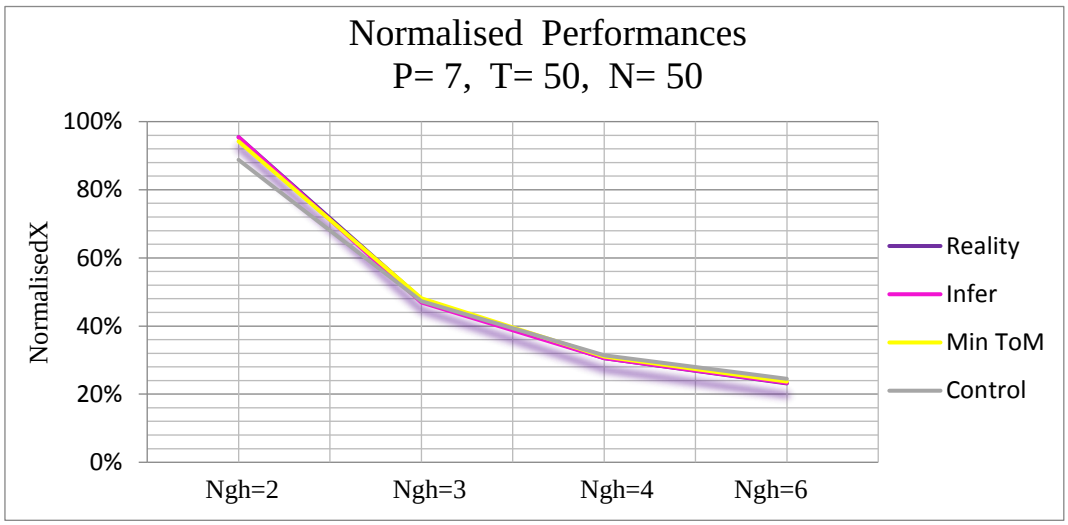


Figure 61. Normalised differences (Infer agents, MinToM agents) with Ngh, (1)

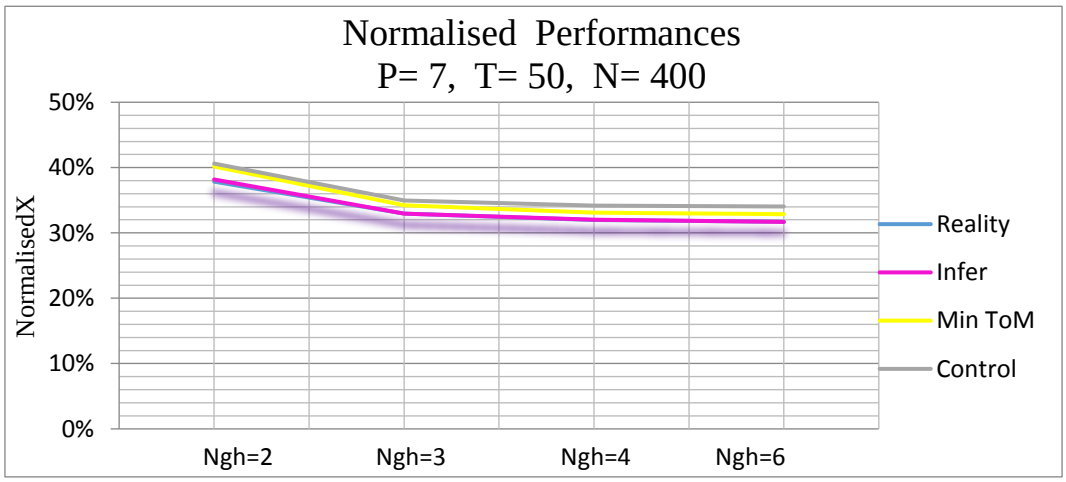


Figure 62. Normalised differences (Infer agents, MinToM agents) with Ngh, (2)

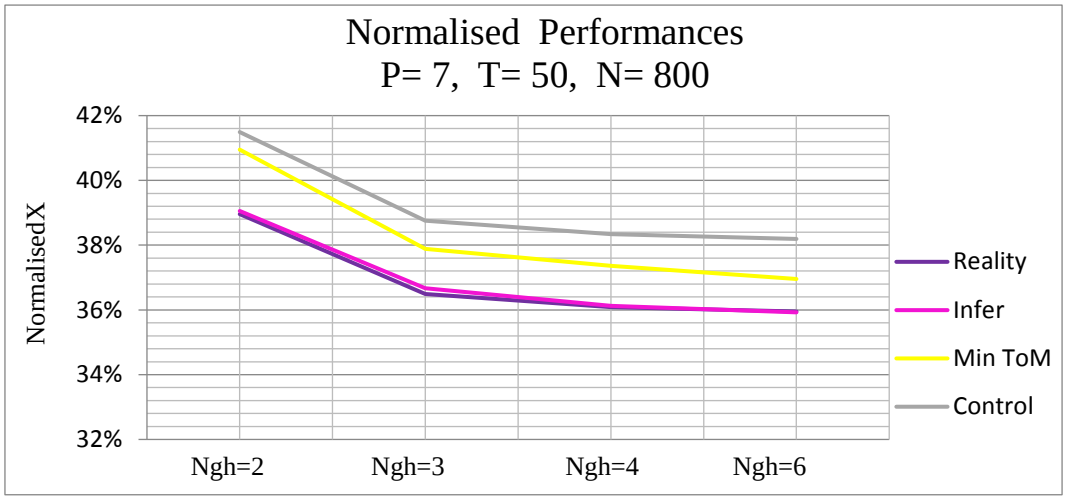


Figure 63. Normalised differences (Infer agents, MinToM agents) with Ngh, (3)

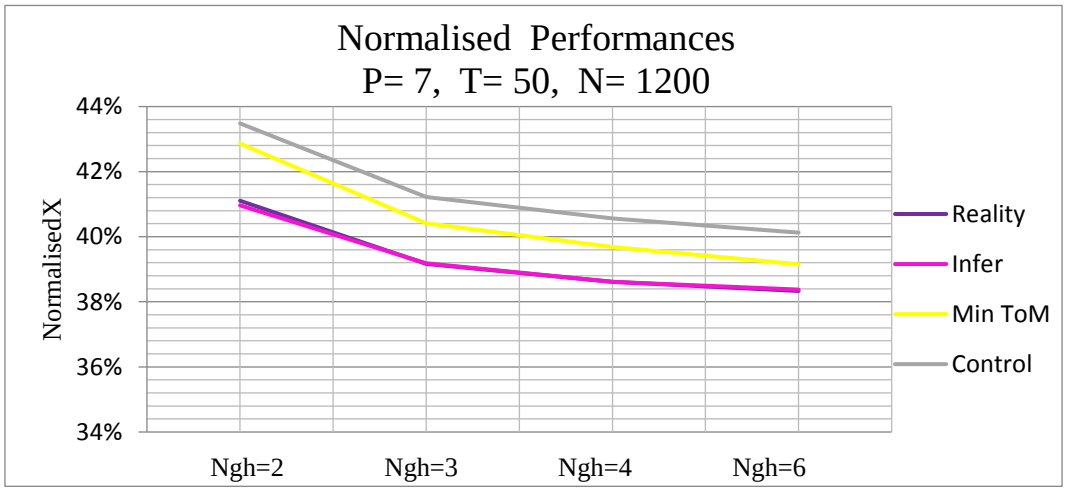


Figure 64. Normalised differences (Infer agents, MinToM agents) with Ngh, (4)

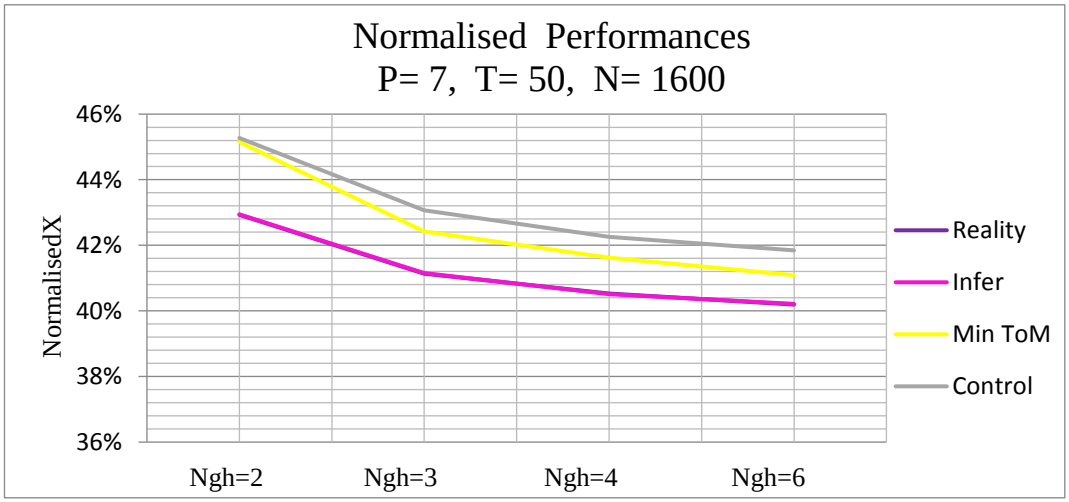


Figure 65. Normalised differences (Infer agents, MinToM agents) with Ngh, (5)

Medium-density Environment

For medium-density environments, the probability of having one food in the first Ngh is high. However, it also depends on the ratio between the number of targets and the number of agents. Thus, the effect of field of view decreases because agents are mainly able to use their individual strategies. In addition, the uncertainty of the environment is lower than for low-density environments. For example, Figure 68 with N=400, T=800 and P=7, shows that the effect of field of view occurs mainly from Ngh=2 to Ngh=3 and the expansion of field of view after that does not have any dramatic effect. This is because it is possible for agents to find targets in the neighbourhood 2 or 3. Thus, agents do not need to use the wider neighbourhood to find targets.

High-density Environment

Noticeably, in situations where the number of targets is high, such that agents can achieve targets without using their abilities, the effect of field of view decreases sharply. For example, Figure 84 with T=1200, N=800, P=2, shows that increasing field of view has little to no effect, especially for Infer agents. The highest level of the differences between Infer and MinToM agents occurs when the ratio between the number of agents and the number of targets is around more than 2/5 and less than 4/5 of an environment of 2500 cells.

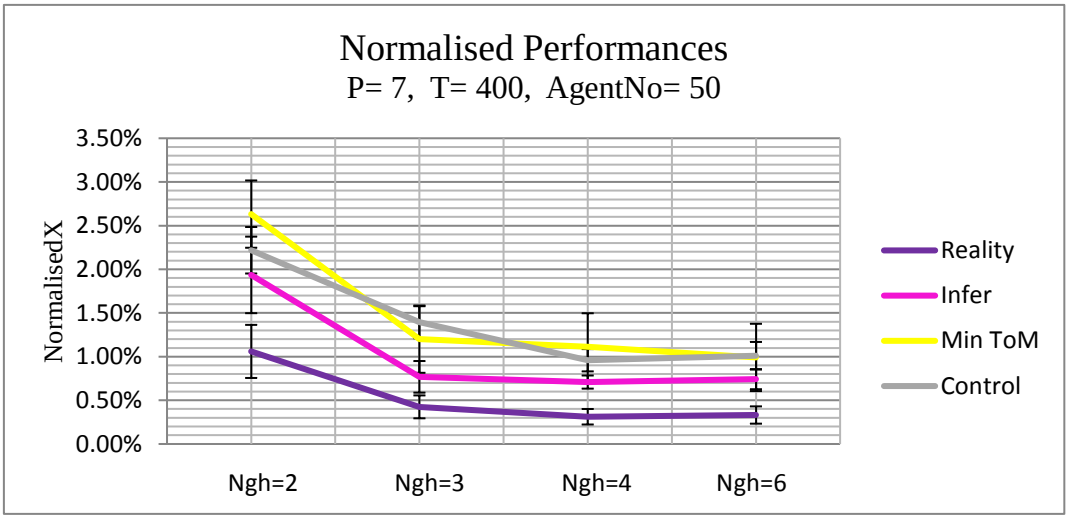


Figure 66. Normalised differences (Infer agents, MinToM agents) with Ngh, (6)

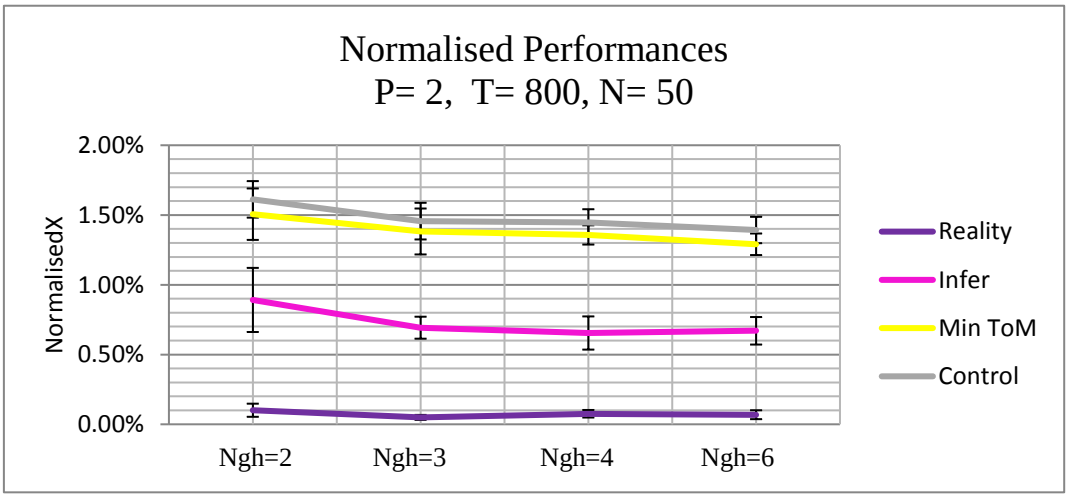


Figure 67. Normalised differences (Infer agents, MinToM agents) with Ngh, (7)

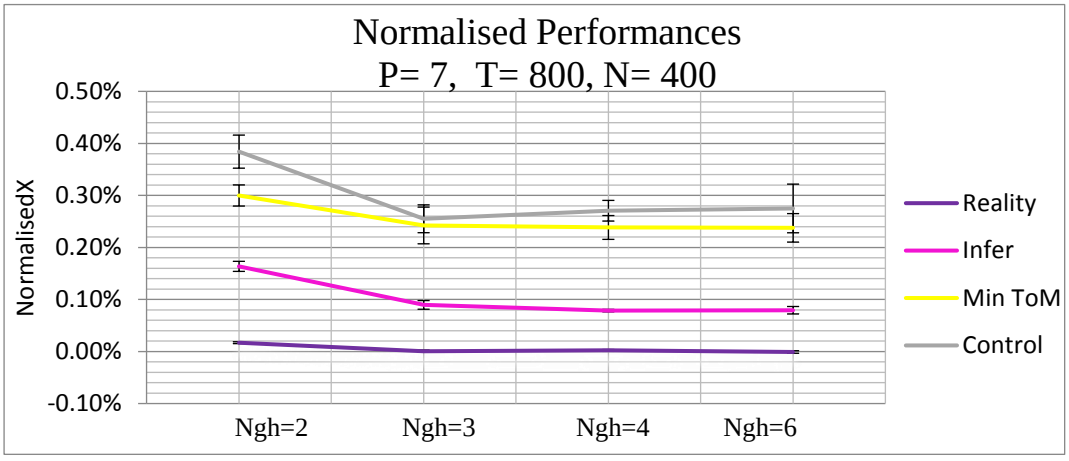


Figure 68. Normalised differences (Infer agents, MinToM agents) with Ngh, (8)

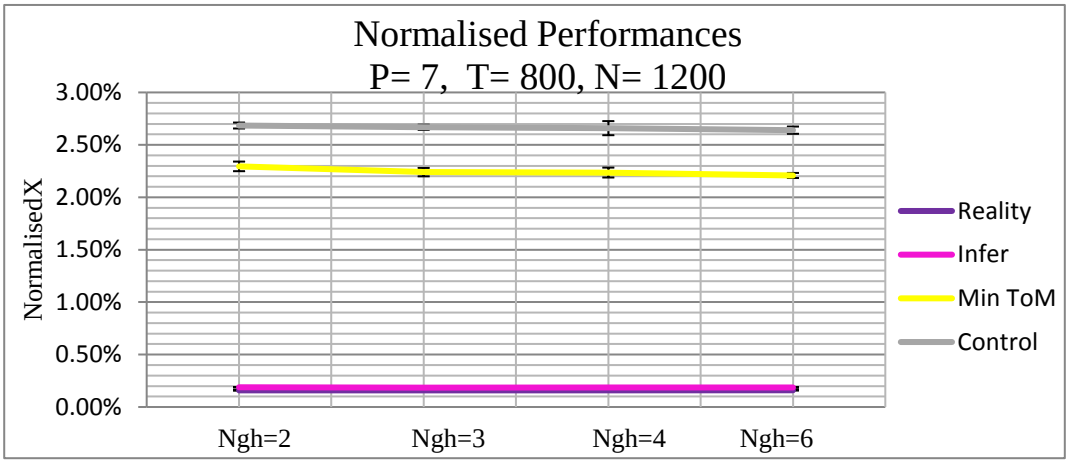


Figure 69. Normalised differences (Infer agents, MinToM agents) with Ngh, (9)

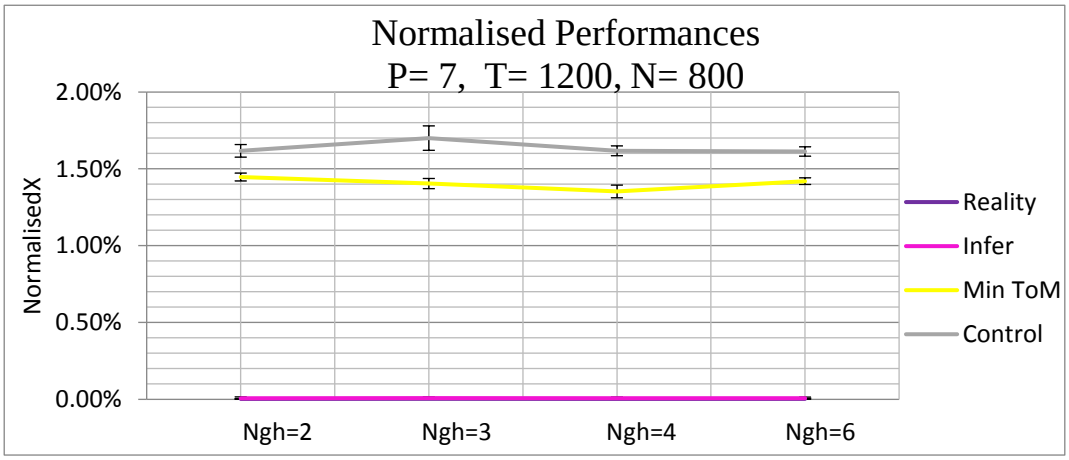


Figure 70. Normalised differences (Infer agents, MinToM agents) with Ngh, (10)

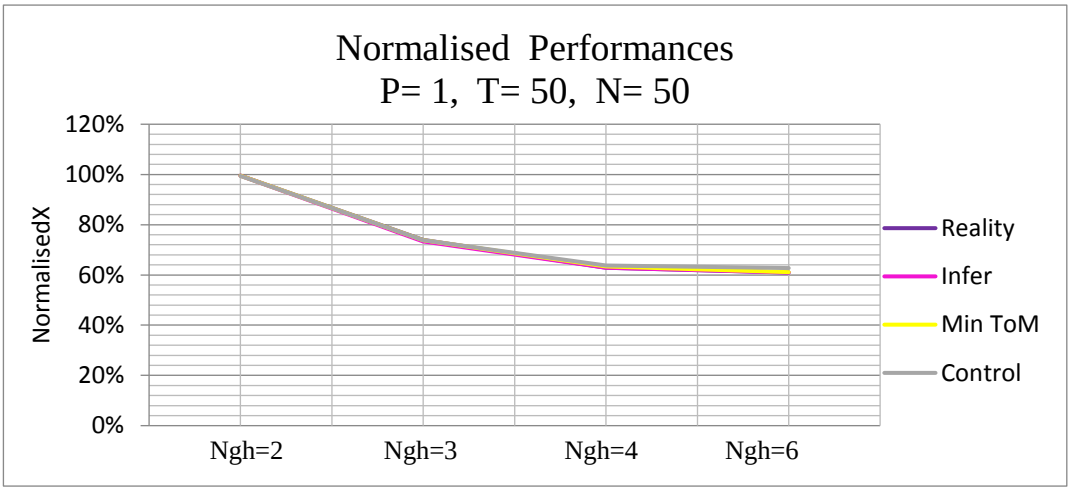


Figure 71. Normalised differences (Infer agents, MinToM agents) with Ngh, (11)

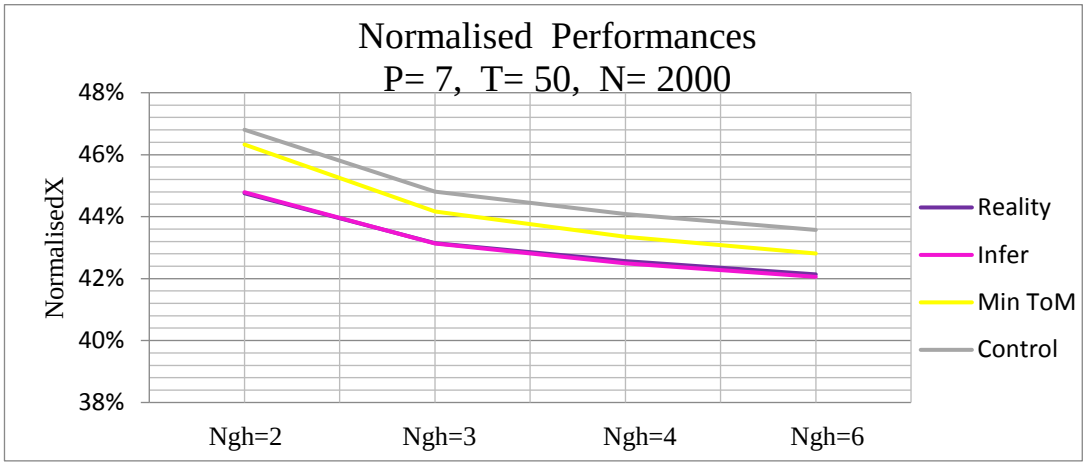


Figure 72. Normalised differences (Infer agents, MinToM agents) with Ngh, (12)

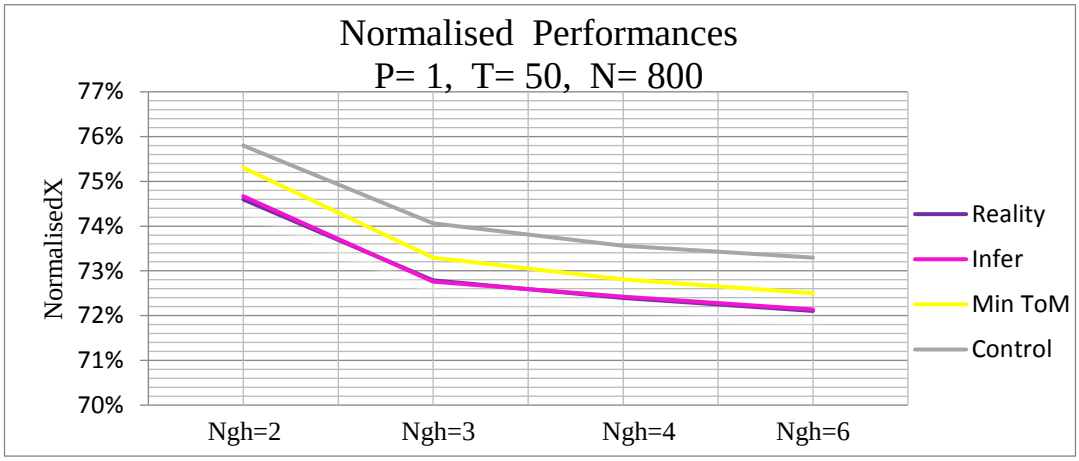


Figure 73. Normalised differences (Infer agents, MinToM agents) with Ngh, (13)

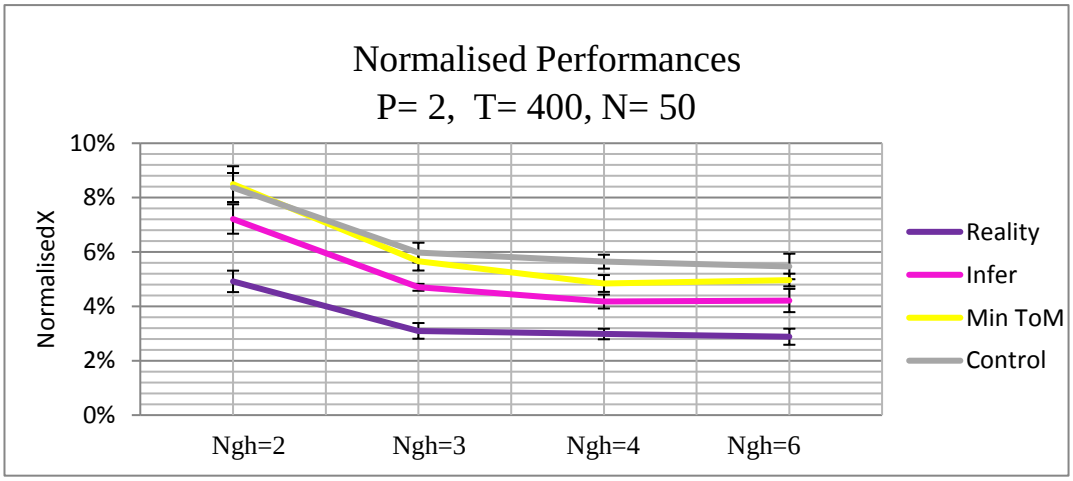


Figure 74. Normalised differences (Infer agents, MinToM agents) with Ngh, (14)

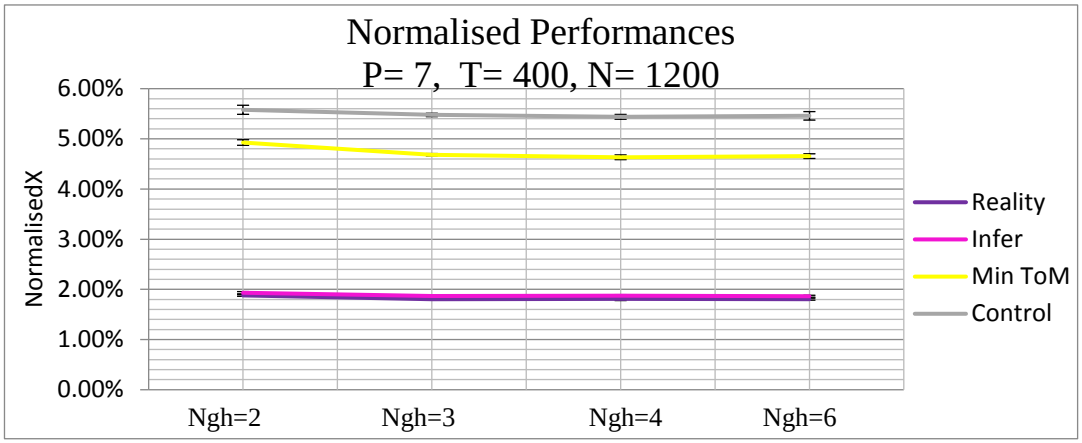


Figure 75. Normalised differences (Infer agents, MinToM agents) with Ngh, (15)

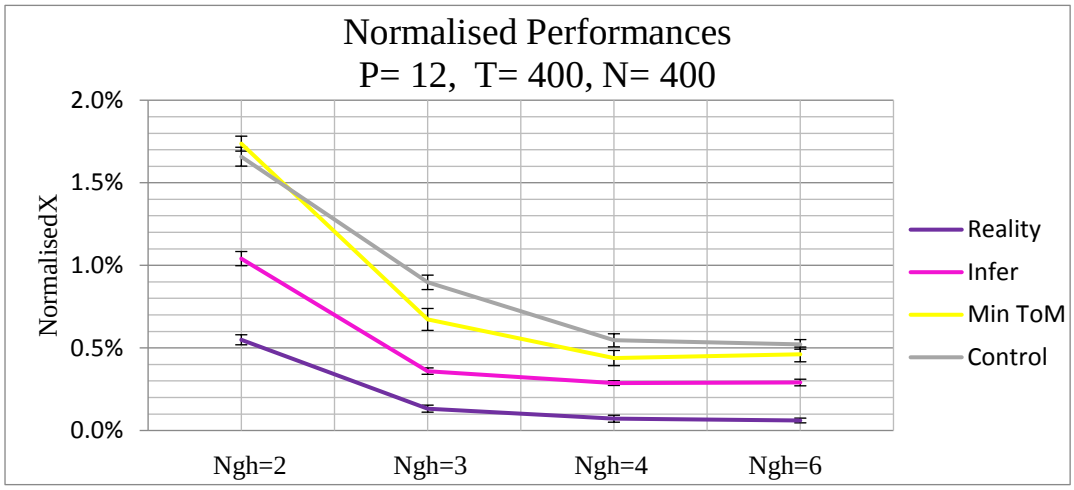


Figure 76. Normalised differences (Infer agents, MinToM agents) with Ngh, (16)

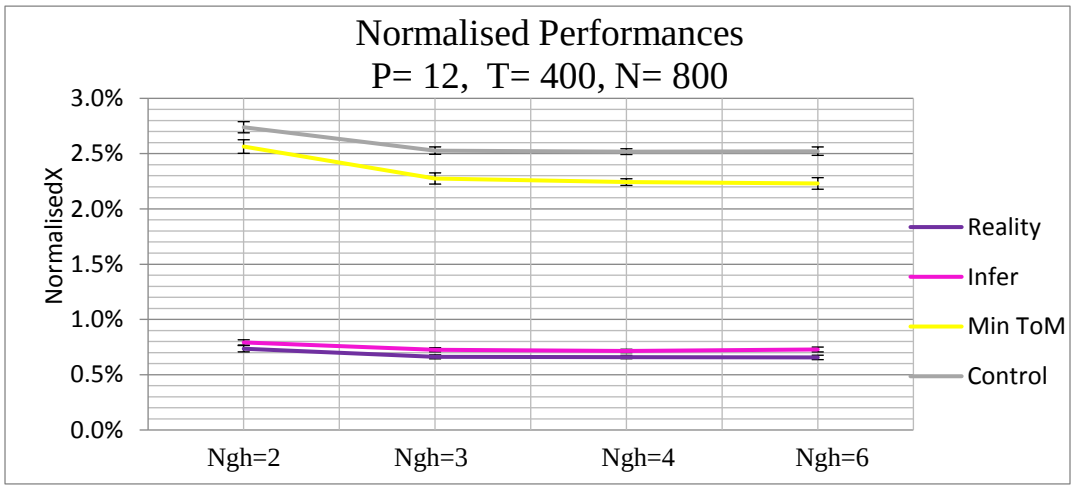


Figure 77. Normalised differences (Infer agents, MinToM agents) with Ngh, (17)

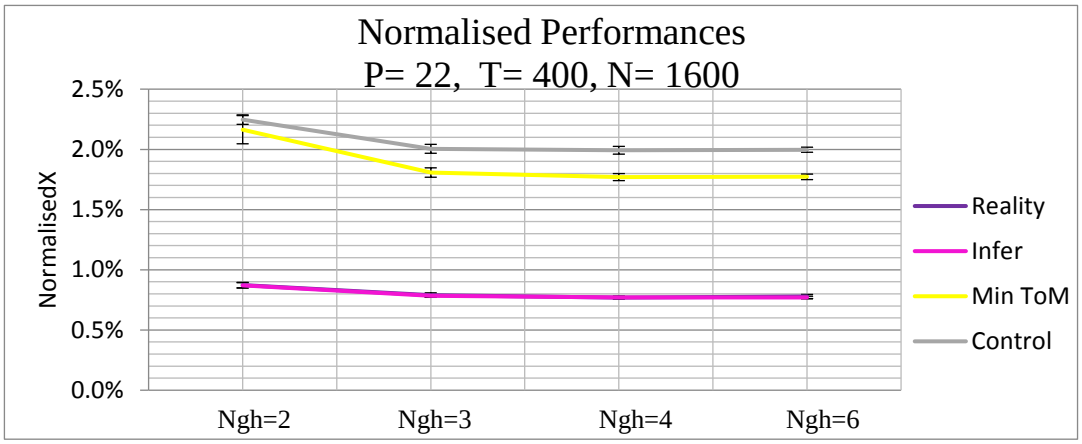


Figure 78. Normalised differences (Infer agents, MinToM agents) with Ngh, (18)

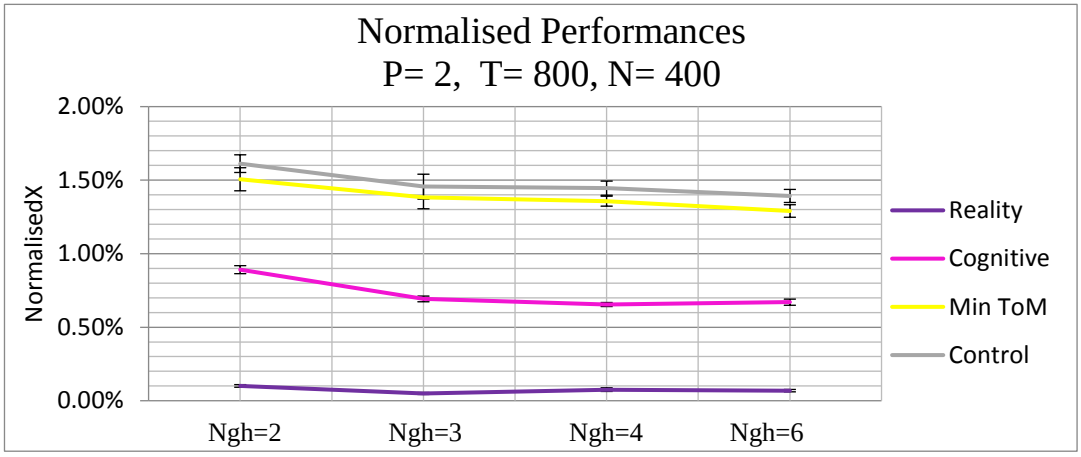


Figure 79. Normalised differences (Infer agents, MinToM agents) with Ngh, (19)

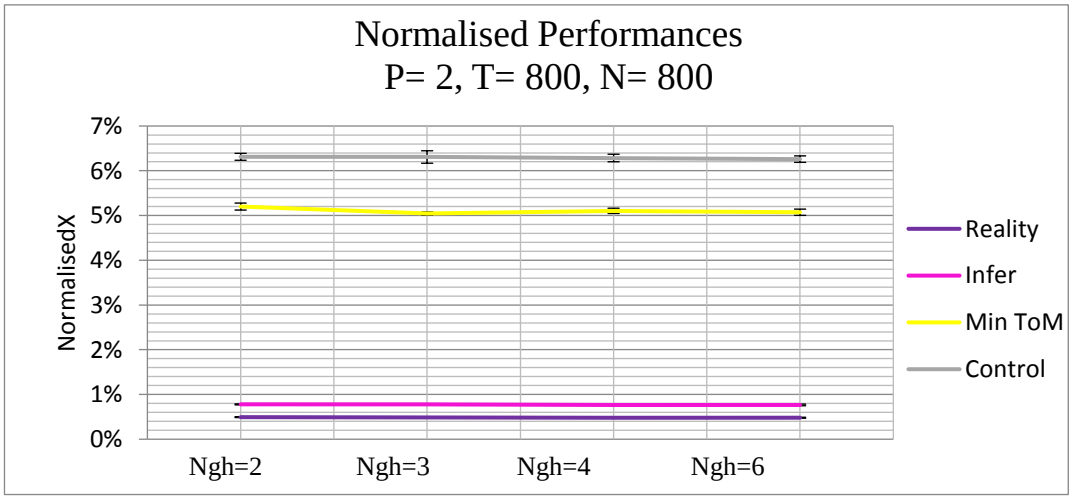


Figure 80. Normalised differences (Infer agents, MinToM agents) with Ngh, (20)

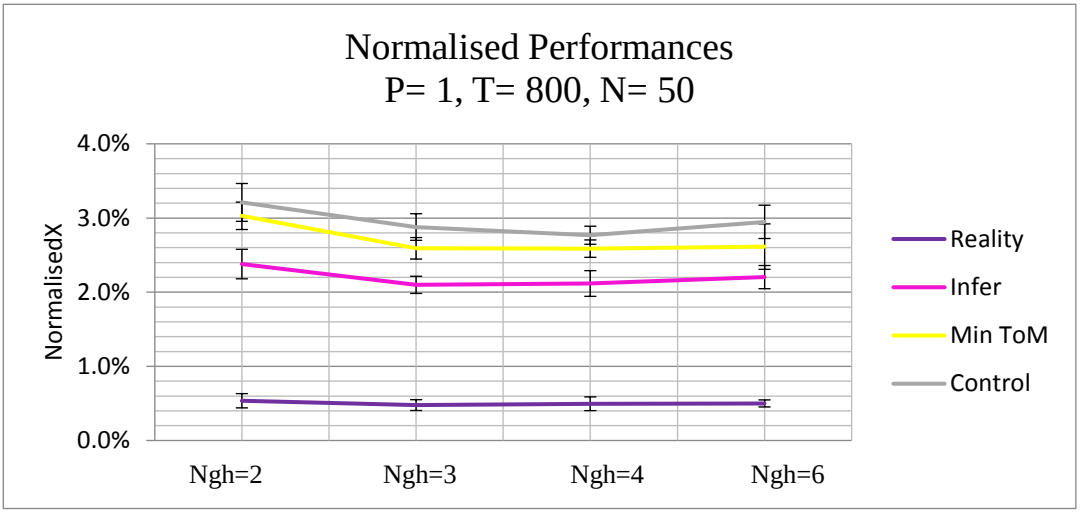


Figure 81. Normalised differences (Infer agents, MinToM agents) with Ngh, (21)

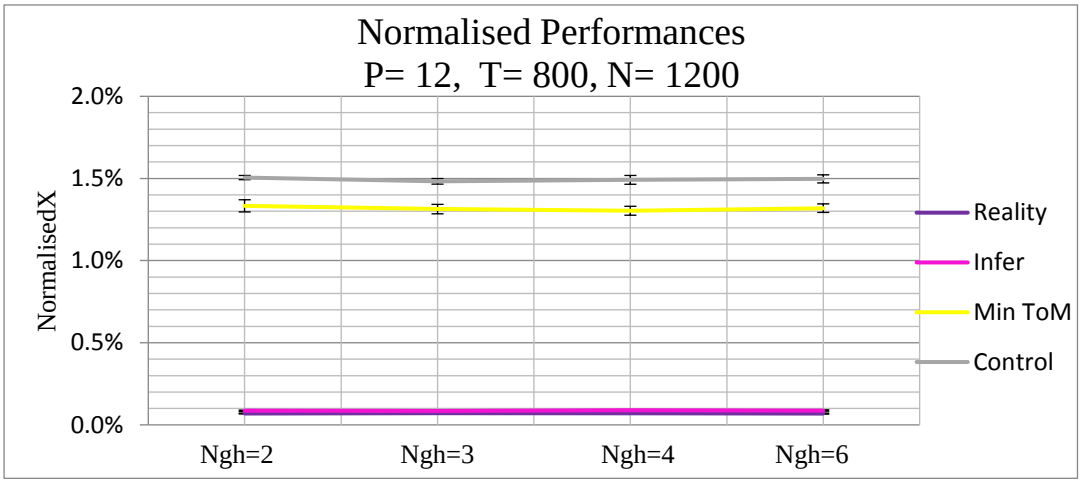


Figure 82. Normalised differences (Infer agents, MinToM agents) with Ngh, (22)

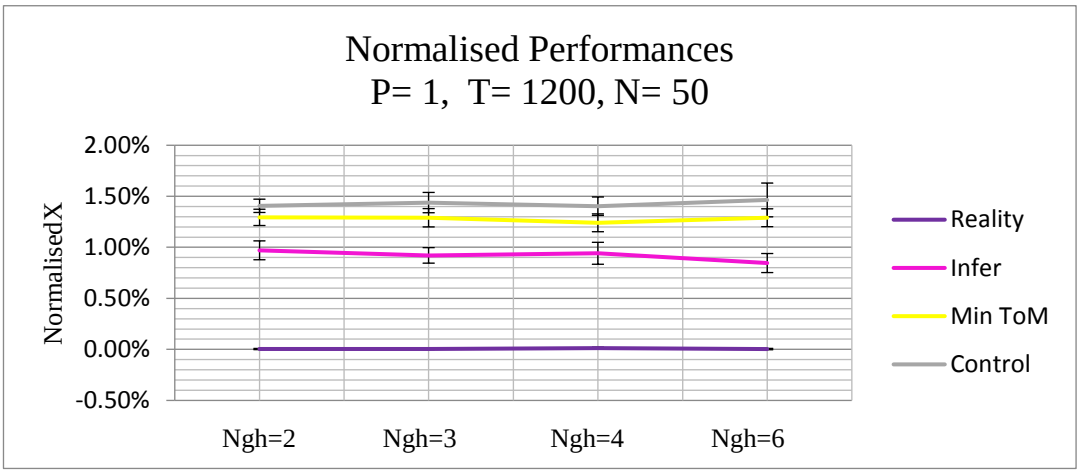


Figure 83. Normalised differences (Infer agents, MinToM agents) with Ngh, (23)

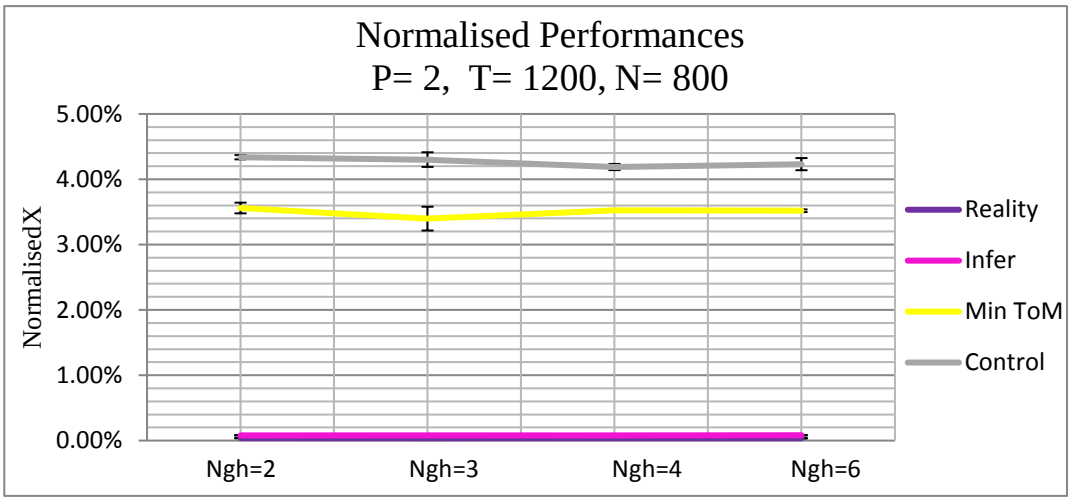


Figure 84. Normalised differences (Infer agents, MinToM agents) with Ngh, (24)

In addition, graphs from Figure 95 to Figure 99, in appendix 1, show how many times Infer agents have used their expanding Ngh function, as well as the number of times that they have used their infer system and thus theory of mind ability.

3.3.3 Differences of Infer and MinToM agents based on maximum and minimum performances

The results of the simulation have shown that the descending order of agents' performances is Reality, Infer, MinToM and Control agent. Besides, the general pattern of results shows that the largest differences occur between Infer and MinToM agents. Moreover, the large amount of data from simulation runs structured by normalisation in a standard scale of graphs demonstrates differences of Infer and MinToM agents' performances. The focus of this section is to use the minimum baseline of performances of agents (Control agents) and the maximum one (Reality agents) in a way that the normalised data demonstrates differences of Infer and MinToM agents' performances more effectively. For this purpose, the NormalisedX Value of Reality agents is used as the maximum and NormalisedX Value of Control agents as the minimum value for rescaling data which indicates the differences between Infer and MinToM in the range of [0, 1].

Therefore, the efficiency of Infer agents in comparison with MinToM agents is calculated based on the efficiency of Reality agents in comparison with Control agents as follows:

(NormalisedX_Infer - NormalisedX_MinToM) / (NormalisedX_Reality - NormalisedX_Control)

This formula evaluates the Infer agents' efficiency in relation to that of the MinToM agent and normalises it with the difference between Reality and Control agents. Thus, these differences contain the area of variation between the upper and lower limits on agents' performance scale, including agents which lack theory of mind ability (Control agents) and agents with precise understanding of others' mental states (Reality agents) with the same parameters in the simulation.

The results of the simulation and this formula are depicted through Figure 85 to Figure 89. Figure 88 shows the extreme values of N=50 and T=50. The normalised formula does not include the extreme values as agents' performances are random. This figure shows the lowest differences and indicates the lowest agents' performances in an uncertain environment. In particular, the fluctuating differences signal a highly unpredictable environment. By increasing the number of agents, as shown in Figure 89, the stability improves therefore the normalisation formula applies for N= 400-2000 and T=50. The instability of the situation is consistent in other graphs including Figures 85 to 87 for N=50 with an improvement due to increased number of targets. Moreover, Figures 85 to 87 show as the period increases, this low efficiency improves; this is because agents are more likely to stay Passive and do not need to achieve targets. As the number of targets increases to 400 in Figure 87, the efficiency of Infer agents increases, especially by increasing the number of agents, the differences reach an average of 80%. This efficiency continues to rise in Figure 86 for the number of agents of 800 with average differences of under 90%. For T=1200 in, the stability improves for N=50 and remains with high differences of above 90%, as is shown in Figure 85. However, the most significant improvements of differences

occur at T=400 and T=800. One possible reason for this is that as the ratio of T/N becomes larger, the possibility that agents achieve targets regardless of their abilities increases. Thus, the difference between agents decreases.

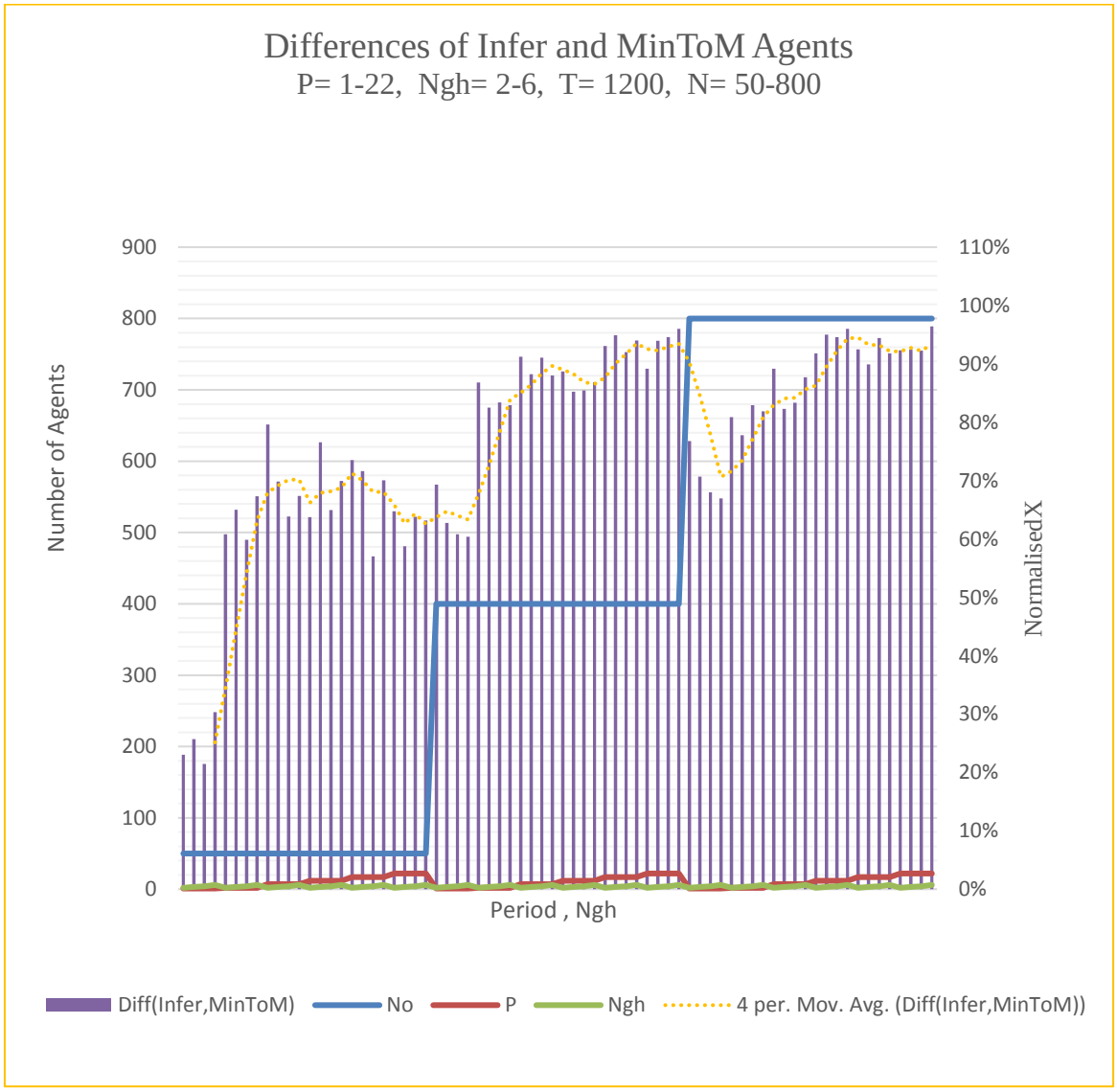


Figure 85. Performance differences of Infer and MinToM agents, T=1200
Setup: Ngh=2, 3, 4, 6 and P=1, 2, 7, 12, 17, 22

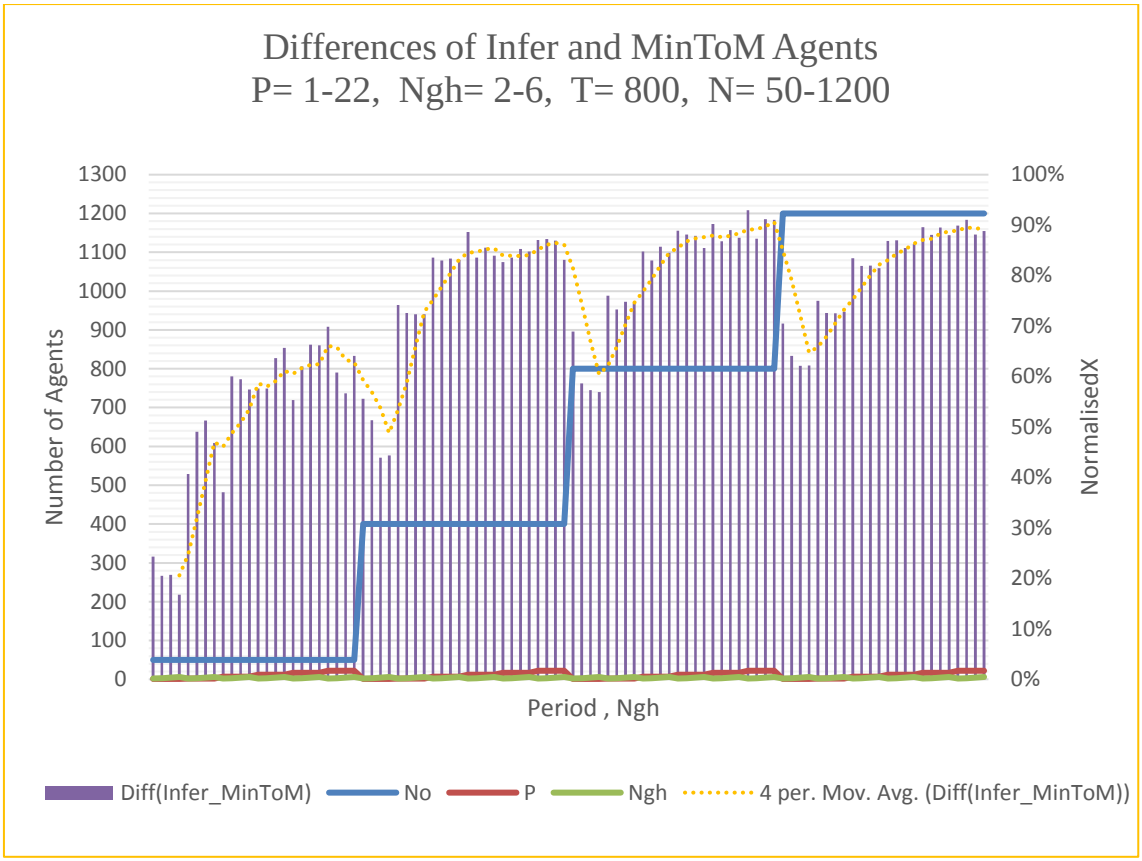


Figure 86. Performance differences of Infer and MinToM agents, T=800
Setup: Ngh=2, 3, 4, 6 and P=1, 2, 7, 12, 17, 22

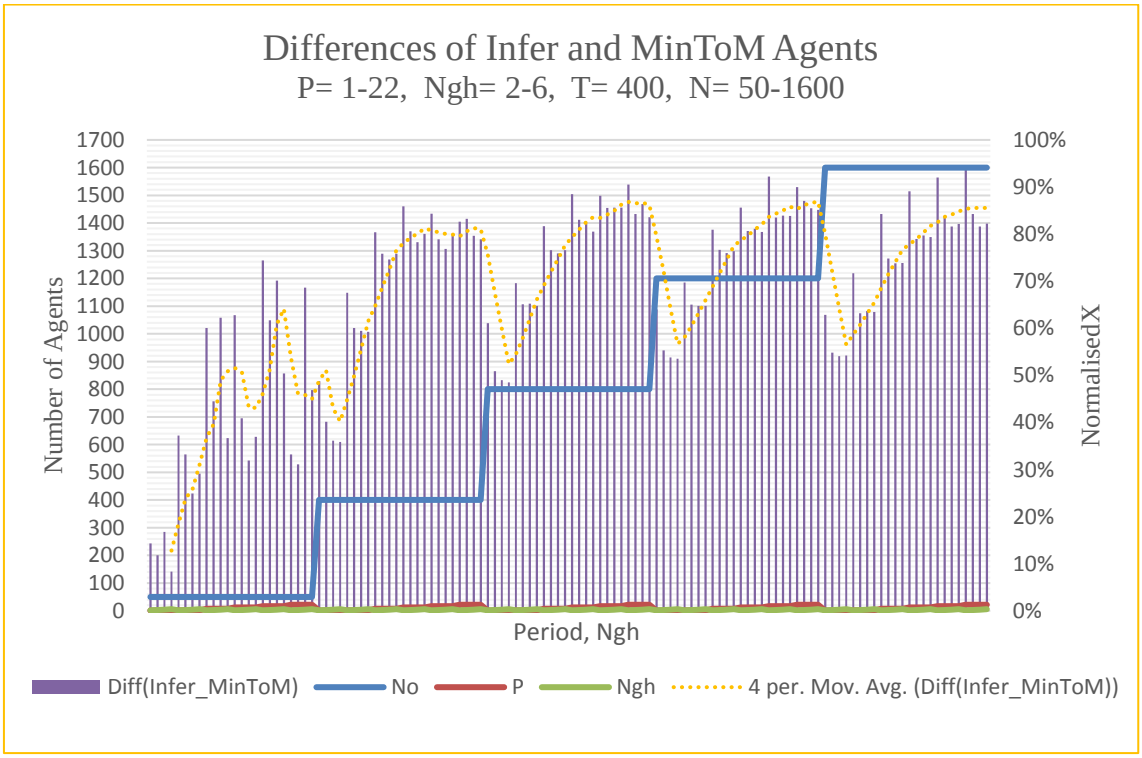


Figure 87. Performance differences of Infer and MinToM agents, T=400
Setup: Ngh=2, 3, 4, 6 and P=1, 2, 7, 12, 17, 22

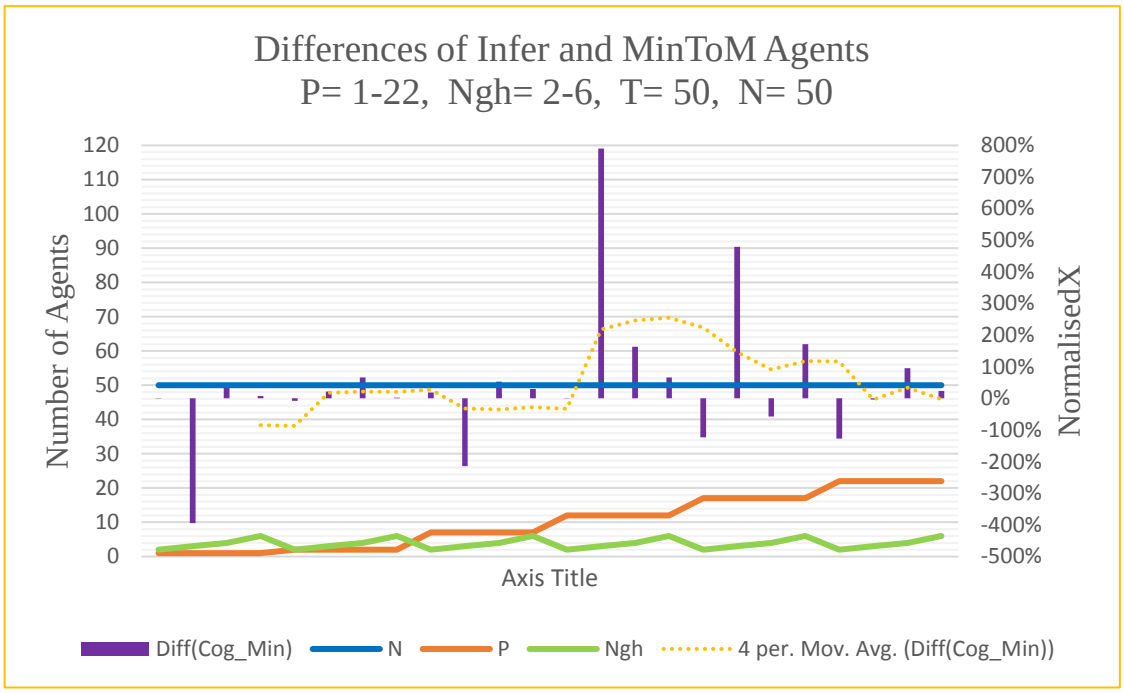


Figure 88. Performance differences of Infer and MinToM agents, T=50, N=50
Setup: Ngh=2, 3, 4, 6 and P=1, 2, 7, 12, 17, 22

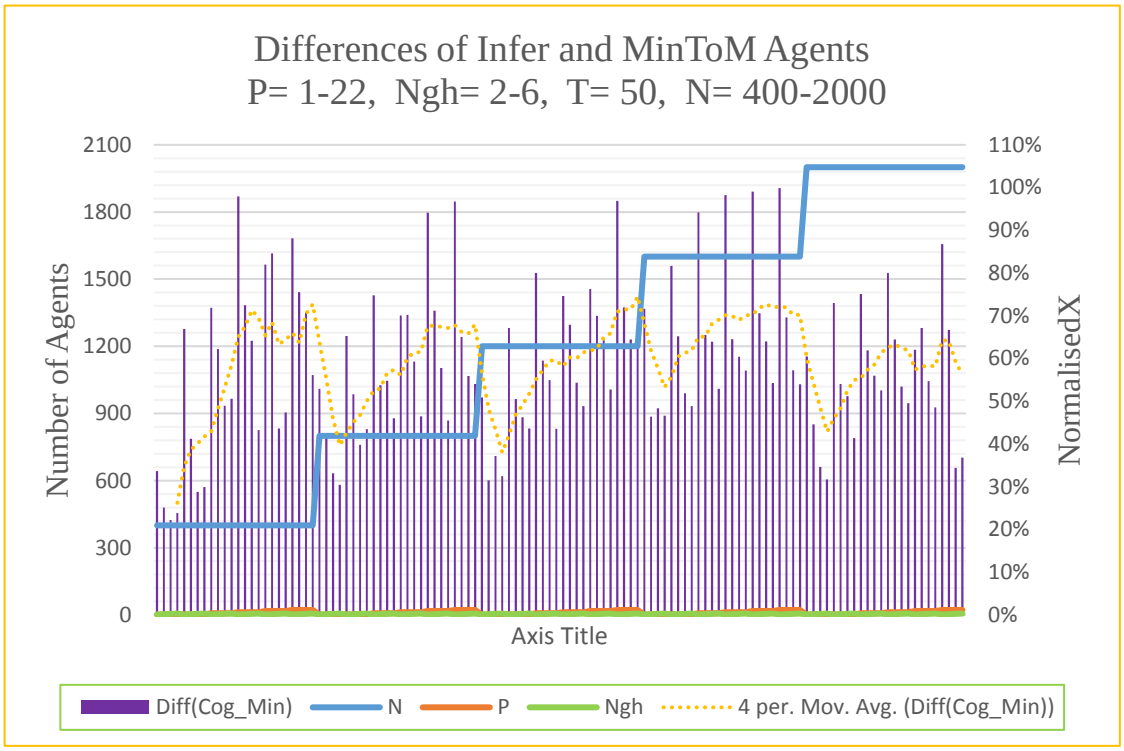


Figure 89. Performance differences of Infer and MinToM agents, T=50, N=400-2000
Setup: Ngh=2, 3, 4, 6 and P=1, 2, 7, 12, 17, 22

Furthermore, the graphs of the simulation illustrating each type of agents' performance in every time step of 1000 ticks are shown in Figure 100 to Figure 105 in the appendix 2. Although, these graphs are beyond the scope of this thesis, the pattern of the graphs show Infer and Reality agents reach their targets faster and stay Passive for the remaining time steps while other agents reach fewer targets and reach an equilibrium faster.

3.3.4 The cost and required resources of inferring others' mental states

Certainly, Infer agents' ability of understanding others' mental states and reasoning demands more resources and costs. The main resources, which have already been described in IAF and BRM, are also applicable in MSM. By referring to Table 4, these resources, comprising a network of perception and attention (first phase of RAF in MSM), three types of memory for storing information (second phase of RAF in MSM), inhibitory control and reasoning resources (third phase of RAF in MSM). Noticeably, other agents such as MinToM and Food agents require a very limited version of these resources, for example, MinToM agents require sensory and short-term memory but they do not need long-term memory, inhibitory control and reasoning resources.

Furthermore, in order to calculate the inferring others' mental states cost, the agents isolated processing time is computed for each type of agent. However, this processing time had no effect on the simulation and MSM results because of software time settings (Repast Simphony). The results demonstrate that Random agents are the fastest and that Infer agents are the slowest. There are two possible reasons for Infer agents requiring long processing time; firstly, reasoning beliefs and desires in both reasoning phase and expressing phase takes additional time to gather and evaluate the corresponding information as the amount of information is larger than other agents. Secondly, inhibition step requires a shift from Infer agents' own beliefs and desires in the present time to others,

which consumes more time for retrieving the information from the memory. Hence, based on RAF, devising an infer system and performing with high efficiency comes at the cost of time, as well as efficient interconnected resources including memory, perception, reasoning and inhibitory control and larger amounts of information.

3.3.5 Applying an alternative strategy for Infer Agents

Infer agents' strategy has been a reliable and effective method to achieve targets. The questions then arise: What if the Infer agents' strategy changes? How do these changes affect the agents' performance? What is the role of the strategy that links the information about understanding others' mental state to an action? In order to address these questions, there are two options: First, a more planned and sophisticated strategy, or second, simply reduce some parts of planning of the current strategy. The first option is not fit to answer the questions because it will increase the Infer agents' efficiency. The second option provides insights into the impact of applying less planning. Thus, the Infer and MinToM agents' strategies have been altered by removing the two situations that agents use in their plan, for the next time steps. The first situation involves the agent in an Active state but there is no target available in the field of movement. The second situation involves the agent in Passive state with more than one cell option to move. Thus, instead of applying their strategies, agents move randomly because in these situations there is no direct access to targets. Consequently, the simulation results with this strategy show that the efficiency of Infer agents decreases and matches with MinToM agents' performances.

For example, the results of conducting eight series of simulations for Infer and MinToM agents with and without a plan in two environments with parameters: T=400, P=2, Ngh=3, N=400 and also T=400, P=2, Ngh=3, N=800 are gathered in Figure 90.

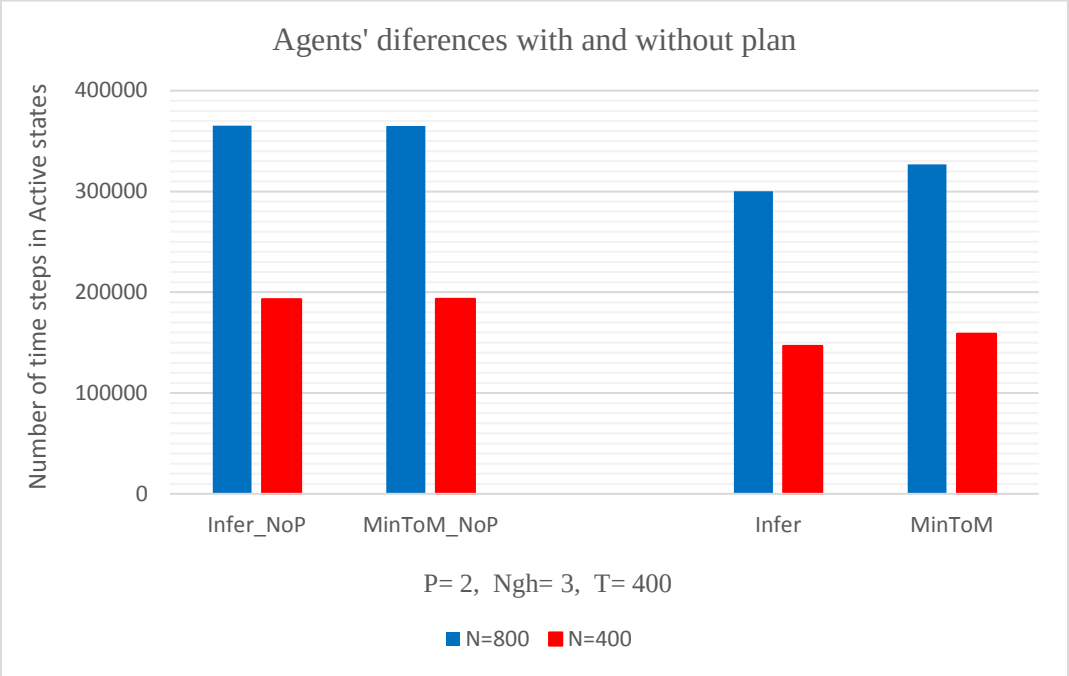


Figure 90. Performance differences of (Infer, MinToM) agents with No plan (_NoP)

This graph shows that Infer agents with No Plan (Infer_NoP) perform similarly to MinToM agents with No Plan (MinToM_NoP) whereas Infer agent performances are more effective in comparison to MinToM agents in the same environment.

Intriguingly, these results suggest that a fruitful use of understanding others' mental states involves a certain level of reasoning and planning. In other words, a successful goal-directed action requires applying the information about others' mental states in rational plans in a competitive society.

3.4 In what way is MSM an effective model?

This thesis presents MSM as an effective model for a simple theory of mind ability. The following explanation for MSM validation and verification provides the model accuracy arguments.

3.4.1 MSM Verification

It is important to describe in depth how MSM works; the agent's simple rules, actions and abilities need to be clear. The descriptions of each agents' strategy and the related diagrams have been extensively discussed in MSM methodology section 3.2. In addition, Table 4 indicates different type of agents' perceptions, functions and memory abilities; all of these conceptual descriptions especially regarding Infer agents and MinToM agents determine how they carry out the intended theory of mind ability and tracking others' field of view respectively. Figure 33 shows the inference scheme, the steps where Infer agents observe other agents achieve targets and update its desires.

The code verification for each agent has been tested to examine that the corresponding code works correctly based on agents' methodology. A code verification cycle including applying a breaking point, running and testing the printed results for each sub function is implemented for each type of agent.

Furthermore, by altering four parameters' values in the initial conditions including: number of food, number of agents, agents' field of view and period of staying Passive, model sensitivity to its initial conditions are examined. Agents' performances are sensitive to the initial conditions, particularly where agents are unable to apply their rules and strategies in uncertain situations. The analyses are in the results section and previously explained.

3.4.2 MSM Validations

MSM validations mainly consist of validating that Infer agents are able to infer others' mental states and that MinToM agents are able to track others' field of view. The micro level validation has already been explained in strategy section. Moreover, the MSM macro level validation has been shown in the simulation results and graphs. The MSM aggregated results and concluding statements are described in the results section.

MSM implementation involved a rigorous modification cycle of the design and programing of different agents. Specifically, Infer agents and MinToM agents have been constantly examined to ensure they theoretically represent the expected methodologies.

3.5 DISCUSSION

The MSM simulation results consistently suggest that there is a strong relationship between the agents' theory of mind capability and their performance efficiency. Theory of mind competence is a reliable factor for higher performance in this competitive society. The reason for Infer agents' high performance is their ability to infer others' desires in their neighbourhood. The second higher performances are MinToM agents due to their ability to track others' field of view. Control agents, which are only able to observe the wider area, have the third most effective performance. In addition, Food agents with understanding of their own desires are in the fourth position whereas Random agents, which move randomly, are the least efficient position. Moreover, Reality agents' best performance is due to their direct access to others' mental states and they are designed for control measurement purpose.

As the ability of theory of mind develops from the simple level of understanding their own mental states to the level of tracking others' field of view and to the more complicated level of constructing inferences about other agents' mental states, the performance of agents effectively increases.

Similarly, de Weerd et al. (2013) assessed whether higher-order theory of mind advances individuals' performances in an agent-based model for a single-shot rock-paper-scissors and Limited Bidding games. Agents with first-order theory of mind can consider the game from the perspective of their opponents, and determine what they do in the position of the opponents. Agents with a second-order theory of mind ability believe that their opponents might use the first-order theory of mind to predict their behaviour. The results of their

simulation shows agents perform better than their rivals that are more limited in their ability to model others; particularly their study demonstrates that agents with the first-order and the second-order theory of mind ability would benefit upon their opponents with lower order in simulation (de Weerd, Verbrugge, & Verheij, 2013). However, their study did not show the same pattern for more than the second order.

These findings are consistent with theory of mind's impact on social situations in real life. For example, the absence of theory of mind (in autism) involves with deficits in understanding and reasoning about mental states (Tager-Flusberg, 2007). Moreover, Volkmar et al. show in a study that autistic individuals exhibit social and communication dysfunction (Volkmar et al., 1987). The study by Frith et al. (1994) found children who lack theory of mind are very low level in social adaptation. The research by (Baron-Cohen , 1989) claims that there is a lack of social abilities in autistic individuals. The majority of research suggests that theory of mind deficit is a potential reason of social dysfunctions (Peterson et al., 2009) and the mainstream of these studies focus on impaired theory of mind ability in children. In fact, the link between theory of mind and social life have been studied by taking different perspectives. For example, Humphrey proposed a social intelligence hypothesis that suggests social competition generally activated primate cognition (Humphrey, 1976). Alternatively, competition for resources in group living, may have caused in primates to evolve 'social intelligence' (Whiten, 2000).

The results of differences between agents' performances in MSM demonstrate a solid pattern with their underlying strategy and the maximum amount of information they need. As it might be expected, the MSM results show that the largest differences happen between Random agent and other agents. This shows that the consequence of having no strategy results in the lowest performance. Indeed, having no strategy signifies there is no need to use any information.

Besides, the simulation results show that the differences between MinToM agents and Infer agents are consistently significant through various values of parameters. This finding indicates that the ability to understand others' mental states plays a critical role in agents' performances in this competitive environment.

The highest level of the differences between Infer and Min ToM agents occurs when the density of the total number of agents and targets is more balanced between 2/5 and 4/5 of the environment cells, indicating the strategy of Infer agents is more applicable in such environments. Almost all of the simulation results show that Infer agents' performances are higher than MinToM. The ability to track others' field of view is used by both of agents. The two main abilities of Infer agents consists of; firstly, their infer system to infer others' desires and secondly, tracking others' field of view. MinToM agents possess the second feature. The second ability might be used with or without the first one. For example, in situations that Infer agent is unable to observe other agents achieving a target, it is consequently unable to update the information and infer others' mental states. Thus, Infer agent can only rely on its own ability to track others' field of view. It applies its pre-assumption and considers all of agents' mental states as Active, similar to MinToM meaning MinToM agents are a basic version of Infer agents. Therefore, MinToM agents' ability is a subset of Infer agents' ones.

Infer agents' complex strategy demands more information, processing and reasoning resources which results to their high level of performance. Whereas, in MinToM agents' strategy, the maximum amount of information they need and their reasoning level are lower than corresponding ones in Infer agents which makes their performance less efficient than Infer agents.

On the other hand, the performance difference between Infer agents and Reality agents are low. Almost similar level performances of Infer and Reality agents indicates that Infer

agents inferences works reliably. It also proposes that inferring other agents' mental states has a significant impact on Infer agents' performances.

Although Reality agents have the highest performance, yet their information is limited to their field of view likewise other agents. In addition, their strategy is not perfect and it still can be improved. Hence, Reality agents' performances are not completely ideal and faultless.

The simulation results show that the lowest differences occur between Food and Control agents. The strategy of Control and Food agents are similar for Active states, as it has already been explained in section 3.2.3. The main difference is that Food agents' field of view is limited to 2, while Control agents' field of view depends on the parameter of field of view. A comparison between these findings and the differences between Infer and MinToM agents concludes that the effect of understanding others' beliefs and desires is significant whereas expanding field of view effect has less effect based on their performances.

In addition, the Infer agents' high performance comes at the cost of time because they need more time (not in the simulation time because of the software's settings) to collect more information and importantly to reason, inhibit their own perspective and retrieve the information regarding others' mental states. Correspondingly, RAF phases are consistent with the developmental literature demonstrate a network of resources for Infer agents (with theory of mind competence) including perception, memory, self-perspective inhibition and reasoning resources.

Finally, Infer agents' strategy has been altered to a version with less reasoning regarding others' mental states. By removing some rational points and links about other agents' mental states (Passive or Active), their performance decreases. This shows the different approach of actions influences agents' performance where using theory of mind

understanding. The result of altering the strategy initially elucidates the significance of reasoning interplay in theory of mind actions. These results suggest that one potential reason for individual differences in their actions (based on theory of mind) is diverse approach of reasoning. Although the information about others' mental states are identical, but the agents' actions require proper rational reasoning to achieve their goals. Noticeably, Hughes has already stated this point that children do not perform their theory of mind understanding similarly in their social life (Hughes, 2011-a).

3.5.1 Infer Agents diagram (RAF)

The process of understanding others' mental state is a complex system and it requires a dynamic analysis. The diagram of Infer agents (RAF) clarifies and verifies the crucial phases of this procedure. It considers the objective of the process and interactions between these phases simultaneously.

RAF simplifies and generalises the main phases of understanding others' mental states process which is shown in Figure 31. RAF reflects the process of how Infer agents infer other agents' beliefs and desires and use this information within the four phases. These phases are as follow:

- Collecting Information

Collecting information is crucial in understanding others' mental states. Infer agents collect information about other agents' perspective from their field of view as well the location of food and other agents. Infer agents observe other agents' actions when they reach a target and track other agents' field of view regarding the locations of targets. There is a reasoning behind which information to collect and use first. For example, they use the closest information first, and when unable to reach their targets, then they expand their field of view to observe more information.

Infer agents observe other agents' actions of reaching a target and use this information to feed the reasoning phase. In addition, they collect information about the location of targets as input for the other phases, enabling them to make decisions where to move. Thus, this phase is highly interconnected with other phases of RAF.

Collection information phase is a dynamic and online procedure, which is parallel with the changes of the world over time.

- Recording Information

Undoubtedly, MSM shows that there are memory demands on Infer agents to understand others' mental states. Infer agents store other agents' desires and beliefs regarding the targets in their memory, which will be accessed in the reasoning phase. Infer agents are able to store the information by exploiting the relevant types of memory. In fact, there are three types of memory. The first type relates to very simple information from the environment such as the location of the food in the current time step (sensory memory). The second type is designed to store simple calculated information about other agents' perspective in the current time step (short-term memory) and the third type stores the inferred information about other agents' desires and beliefs from past time steps (long-term memory). These types of memories are essential for Infer agents and enable them to have access to necessary information especially information about others' mental states. The two main distinctions between these types of memory is associated, firstly, with the length of time that information remains in the memory and secondly with the volume and complexity of their content especially with the information about others' perspectives.

- Reasoning Process of Beliefs and Desires

This phase defines a central information processing unit for Infer agents' theory of mind ability. Infer agents possess their own beliefs and desires. Based on these desires and beliefs, Infer agents choose the agents in their field of view which have access to shared

targets. Then the Infer agents detect and infer the chosen agents' beliefs and desires. Infer agents inhibit their own desires towards the target and temporally retrieve the other agents' perspective from the memory. Once Infer agent observes that another agent, for example agent Z reaches a target, it concludes that agent Z's mental state was Active and thus agent Z reached a target. Therefore, their mental state is Passive in the current time step. This enables Infer agents to infer others' desires. Infer agents track other agents' field of view, which have the same belief about the location of a shared target. Thus, they initially choose the agents with the same beliefs about a target and infer their desires. In general, there are four different mental states situations for both of the Infer agent and the other agent (agent Z), which is shown in Table 6. Hence, the reasoning phase procedure rests on analysing these mental states. Infer agents' actions are different for each of these four cases by considering first their own and then others' mental states.

For example, if Infer agent is in an Active state and infers that agent Z's mental state is Active, then the Infer agent assumes that agent Z believes that there is a target at a specific location and therefore, Infer agent considers that agent Z believes that this is a potential target. Otherwise, if agent Z's mental state is Passive, depending on Infer agent's own mental state it might consider calculating the number of remaining time steps until agent Z becomes Active. In sum, Infer agents assess whether it is necessary to consider other agents' mental states. This assessment is contingent on four different conditions of its own and other agents' mental states which was described earlier.

Agent	Other		
	Mental state	Active	Passive
Infer	Active	Active, Active	Active, Passive
	Passive	Passive, Active	Passive, Passive

Table 6. Four mental state cases between Infer agent and other agents

Reasoning phase for Infer agents involves the following five generic subroutines:

- Inferring others' desires from their observed behaviour
- Processing its own desire and belief
- Self-perspective inhibition of its own desire
- Retrieving information about the subject agents' mental states (from memory)
- Subject agents' desire and belief process

By the end of this phase, Infer agents' reasoning about the other agents' mental states are completed. However, Infer agents have not revealed the understanding of other agents' mental states in their actions yet.

- Expressing Others Mental States (Actions as output)

This phase of the MSM is concerned with deciding on actions by considering others' mental states. Infer agents' actions are influenced by others' desire and beliefs and also the current state of the environment. Therefore, Infer agents' action is based upon two types of information; firstly the information which they directly perceive from the environment and secondly the information which they infer about others' mental states. There is a delicate but important distinction between understanding others' mental states and using this understanding of others' mental states in Infer agents' actions. This clarifies the logic behind the distinction between recognising others' mental states in the first three RAF phases and the expressing phase. In two identical situations of the environment, Infer agents' actions depend on other agents' mental states; they act differently based on whether the other agent's mental state is Active or Passive. Infer agents use the information resulting from their theory of mind ability in this phase through their actions. A level of inevitable reasoning in this phase rests on principles of rational action to achieve their goals. This level of reasoning combines the concept of understanding others' mental states through an action which expresses the understanding of others' mental states.

3.5.2 RAF and MinToM agents' process

The processes of MinToM agents is not exactly the same as Infer agents in RAF as one might expect. However, there are similarities between them. In the first phase, they collect information about others' field of view. In phase two, they register the information and use this in the current time step but do not record this information for future time steps. MinToM agents do not need the reasoning phase of RAF because they only track others' field of view which is not involved with the reasoning phase. Therefore, RAF includes the process of MinToM agents' ability. In sum, the distinction between RAF and MinToM agents' process includes the reasoning phase and the differences between registration and recording as explained earlier.

3.5.3 Infer system complexity

RAF and MSM shows that infer system is an underpinning feature of cognitive theory of mind ability. The relationship between the infer system, the reasoning level and perceived information, in terms of their quantity and quality, determines the level of complexity of the task. As perceived information becomes more complicated, the more robust infer system is required to disentangle others' mental states.

One of the formal distinctions between the first and second system of the two systems of Apperly and Butterfill (2009) concentrates on inference and reasoning capability. The second system relies on the infer system whereas the first system lacks any inferences about others' mental states. Furthermore, the second system information encompasses propositional attitudes as such, whereas the first system involves with simple belief-like such as the location of an object**.** Thus, the two systems implicitly demonstrate the relationship between the perceived information and the complexity level of inferences about others' mental states, which is consistent with MSM results.

3.5.4 Perception, functions and memory in MSM

All of the agents, except Random agents, are able to perceive targets in their field of view. Random, Food and Control agents are unable to reflect other agents' presence and perspective into their plans and actions. Moreover, Control, MinToM, Infer and Reality agents are capable of extending their vision to a larger area for their plans based on the Ngh parameter.

Internal functions including tracking other's field of view, considering their own mental states and understanding others' mental states are main functions to differentiate between the levels of theory of mind ability. The results of these functions provide information that facilitate making decisions for goal directed actions.

As the agents' abilities develop from simple to complicated one on the micro level, it is necessary that agents develop from a simple sensory memory to more advanced storage such as short-term memory and long-term memory. This storage is vital for coding and decoding the required information about inferences. This suggests that agents need different types of memory for different levels of belief representation. MinToM agents, which register the location of the target and are able to track others' field of view, need sensory memory and short-term memory for the current time step. However, Infer agents, with theory of mind ability, need an additional long-term memory to store others' desires. Therefore, the memory of agents become more complicated as the level of theory of mind improves.

3.5.5 Presumptions of mental states (biases)

All types of agents which are Active, except Random agents, have the desire to achieve a target. In addition, they have their presumption about others' mental states. These presumptions are made when agents lack any reasons to think in a different way, these default assumptions (or biases) constitute the most suitable and productive way to make generalisation and often they are not correct (Minsky, 1988).

These presumptions about mental states are updated when corresponding information becomes accessible within the changes of the environment. Infer agents are capable of updating their presumptions regarding others' mental states as they observe other agents achieving targets. The efficiency of Infer agents' decision-making advances through this capability, for example they can move towards a less vulnerable target. At the start of the simulation, Infer agents' presumption is that other agents' mental states are Active which is correct. As other agents' mental states change, this presumption is not necessarily accurate any more. However, the presumptions will be updated with Infer agent's online inferences about other agents' mental states over time and build a reliable input for Infer agents' actions.

Correspondingly, the Reality agent has access to others' mental states as a control measurement and there is no need for mental states presumption.

Furthermore, MinToM agents consider other agents as Active (as their presumption). In contrast with Infer agents, they are unable to update this presumption through the simulation.

Moreover, Control and Food agents assume other agents are always Passive through the simulation. Similarly, they are unable to change this presumption. This presumption, that all agents are Passive, reflects that Control and Food agents ignore other agents. They are egocentric and are only concerned with their own beliefs and desires. The Random agent has no presumption. Table 7 shows agents' presumptions regarding other agents.

Agents	presumption	Others mental states' pre assumption	The ability to update the pre assumption
Random	-	-	-
Food	✓	Passive	-
Control	✓	Passive	-
MinToM	✓	Active	-
Infer	✓	Active	✓
Reality	-	-	-

Table 7. Mental state presumptions regarding other agents

3.5.6 Agents' own and others' mental states

MSM simulate three different levels of mental states understanding for agents, which is depicted in Figure 91. Firstly, Control agents only consider their own mental states and ignore other agents; they are self-centred agents. Secondly, MinToM agents always assume other agents are Active; other agents are considered as competitors with a desire towards the target and they register their beliefs, they are able to track other agents' beliefs about the location of the target. Thirdly, Infer agents have the ability to infer others' desire and register their beliefs regarding the targets. They are constantly updating their perspective about others' desires and beliefs throughout the simulation.

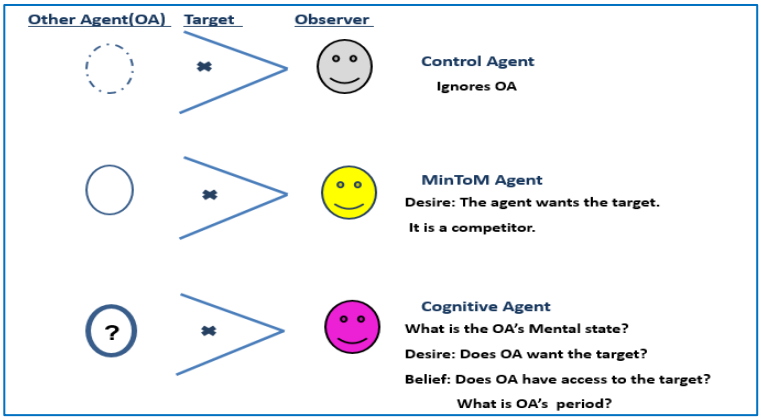


Figure 91. Control, MinToM and Infer agents' understanding of mental states

3.6 Conclusion

One important finding of MSM shows the basic phases of understanding others' mental states processes. The process of how agents infer other agents' beliefs and desires and use this information is reflected within four phases.

In the first phase, collecting information, agents perceive other agents' actions of reaching targets and track other agents' field of view regarding the locations of targets. The primary function of this phase is to provide this information to the other interconnected phases.

Recording phase is the second phase when agents store other agents' mental states in their memory. MSM suggests that there are three types of memory in a simple theory of mind process. Sensory memory relates to simple information from the environment such as the location of the food in the current time step. Short-term memory stores simple calculated information about other agents' perspective in the current time step whereas long-term memory is capable of storing the inferred information about other agents' mental states for future use. The two main features of different types of memory corresponds with the length of time that information remains in the memory and the volume and complexity of the information about others' perspectives. MSM also suggests that as the level of theory of mind develops more complex memory is demanded.

The third phase, reasoning process of beliefs and desires, defines a central information processing unit for agents' with theory of mind ability. This phase involves five generic subroutines; inferring others' desires from their observed behaviour, processing its own desire and belief, self-perspective inhibition, retrieving information from memory regarding others' mental states and finally processing others' mental states.

The fourth phase, expressing others' mental states, is associated with action based upon consideration of others' mental states. The important distinction between understanding

others' mental states and using this understanding in actions is clarified through the first three phases and the expressing phase respectively.

The other MSM simulation results consistently show that there is a solid relationship between the agents' theory of mind ability and the agents' performances. Theory of mind competence is a main factor for higher performance in this competitive society. The reason for agents' high performance is their ability to infer others' mental states in their neighbourhood. The second highest performance belongs to agents with ability to track others' field of view. Agents with no ability of theory of mind, as we would expect, have the third most effective performance. Agents' performance increases as the level of theory of mind develops through the three levels; understanding their own mental states, minimal theory of mind ability, making inferences about other agents' mental states.

Moreover, these four basic phases of theory of mind ability demonstrate that agents with theory of mind competence need a network of resources including perception, memory, self-perspective inhibition and reasoning. This finding is consistent with the developmental literature. The MSM results demonstrate a solid pattern between agents ToM ability level and the maximum amount of information they require. Thus, agents' high performance comes at the cost of time because they need to collect more information and more importantly to reason, inhibit their own perspective and retrieve the information regarding others' mental states.

CHAPTER 4

4. A SYSTEMATIC APPROACH TO BELIEF REPRESENTATION

4.1 Introduction

Recently, theory of mind research has been significantly increased through various fields. However, its underlying processes are still under considerable debate. The lack of structured and standardised basic building blocks for a simple theory of mind in the literature brings confusion in the measurements of theory of mind abilities. Schaafsma et al. (2015) argue that a scientific concept of theory of mind requires a set of simpler processes rather than its current definition as the essence of a mental representation of minds, which does not permit an easy breakdown into its basic components. Consequently, they suggest the reconstruction of a concept of theory of mind with the necessary links to its more basic processes. To achieve this, Schaafsma and her colleagues propose two steps: breaking down theory of mind and its associated concepts into cognitive components that describe more basic processes, and then reassembling these basic blocks into different features of theory of mind.

Accordingly, this thesis offers a generic scheme for many belief-reasoning tasks called Belief Representation Systematic Approach (BRSA), which highlights the key phases of belief representation processes. BRSA is a simple and robust approach that breaks down theory of mind tasks, including false belief tasks, into four cognitive phases describing the basic processes of understanding others' mental states. In addition, BRSA is capable of reconstructing various levels of theory of mind by assembling some features of its phases in different ways.

This chapter provides a standard theoretical framework for theory of mind processes consistent with IAF and RAF from the previous chapters. The aim of this chapter is to develop BRSA and its concept in further detail, explaining where BRSA comes from, BRSA phases in the standard false belief task and BRSA phases in developmental literature. Subsequently, BRSA's explanation for participants' failure in false belief tasks,

as well as the link between minimal theory of mind and BRSA are discussed. Other applications of BRSA, such as analysing the conditions in which the false belief task is a decisive test for theory of mind based on BRSA and in the literature will also be explained. Finally, analysis of minimal theory of mind as a decisive construct for theory of mind based on BRSA will be explored.

4.2 Where does BRSA come from?

Agent-based models embrace a bottom-up principle where all rules and parameters are defined at the micro level (Salamon, 2011). The processes of agents' actions and transitions that originate from the micro level can be shown in a generic diagram. A diagram such as a flowchart is a symbolic description of the model that shows the decisions points and the processes in the computer program (Wilensky & Rand, 2015). "These symbols provide a clear way of understanding how control flows through the software" (Wilensky & Rand, 2015, p. 314).

"The behaviour of agents in many agents-based models is often expressed as logical rules. Hence, it is very common to see flowcharts and pseudo-code (or even computer code itself) used to represent agent behaviours" (Onggo & Karpat, 2011, p. 674). There are visual tools in some platforms that facilitate illustrating agents' complex behaviour diagrams. Some of commercial agent-based platforms include the process flow diagram, functions and behaviour of agents. For example, AnyLogic includes different types of diagrams; Business Process Model Notation (BPMN) is a visual modelling tool for process description and process execution using diagrams for simulation and Repast Simphony includes Statecharts framework, which clarifies the underlying logic of the model.

Therefore, one of the advantages of simple agent-based models is to allow the user to explore the crucial processes in the model comprehensively, and also the effect of these processes on the results (Wilensky & Rand, 2015). The diagrams of agents with minimal

theory of mind ability (Figure 30), agents with understanding of others' false beliefs (Figure 15), and agents with theory of mind ability (Figure 31), are gathered in Figure 92.a to Figure 92.c to show a comparison of agents' procedures with different levels of understanding others' minds. This figure shows that for a minimal theory of mind, three phases of collecting information, registering information and action are essential, whereas for belief representation, the phases of recording information and reasoning other' mental states should be added. Note that in registration, the access to the information is only possible at the current time step while in recording the information, it is possible to store the information for using in future time steps.

The diagram of agents with the ability of inferring others' mental states needs an additional step of inferring others' mental states in the reasoning phase, in comparison to the diagram of agents with belief representation ability. However, the basic phases are identical. This figure illustrates that as the level of theory of mind increases the demand of resources and reasoning rises.

First, BRSA is derived from the diagram of Infant agents (IAF) in BRM, an agent-based model introduced in chapter 2, which illustrates the procedure that occurs within agents with the ability of understanding others' beliefs. This diagram was identified through examining the behaviour and the dynamic processes of decision trees of Infant agents. Thus, the diagram of BRSA is identical to agents with belief representation ability in Figure 92.b, by concentrating in cognitive aspects rather than computational, which is shown in Figure 93.

Essentially, BRSA provides insight into the cognitive processes that drive the complex procedure of understanding others' beliefs. Compatibly, BRSA is applicable for agents with inferring others' mental states ability as a fundamental structure for a simple theory of mind.

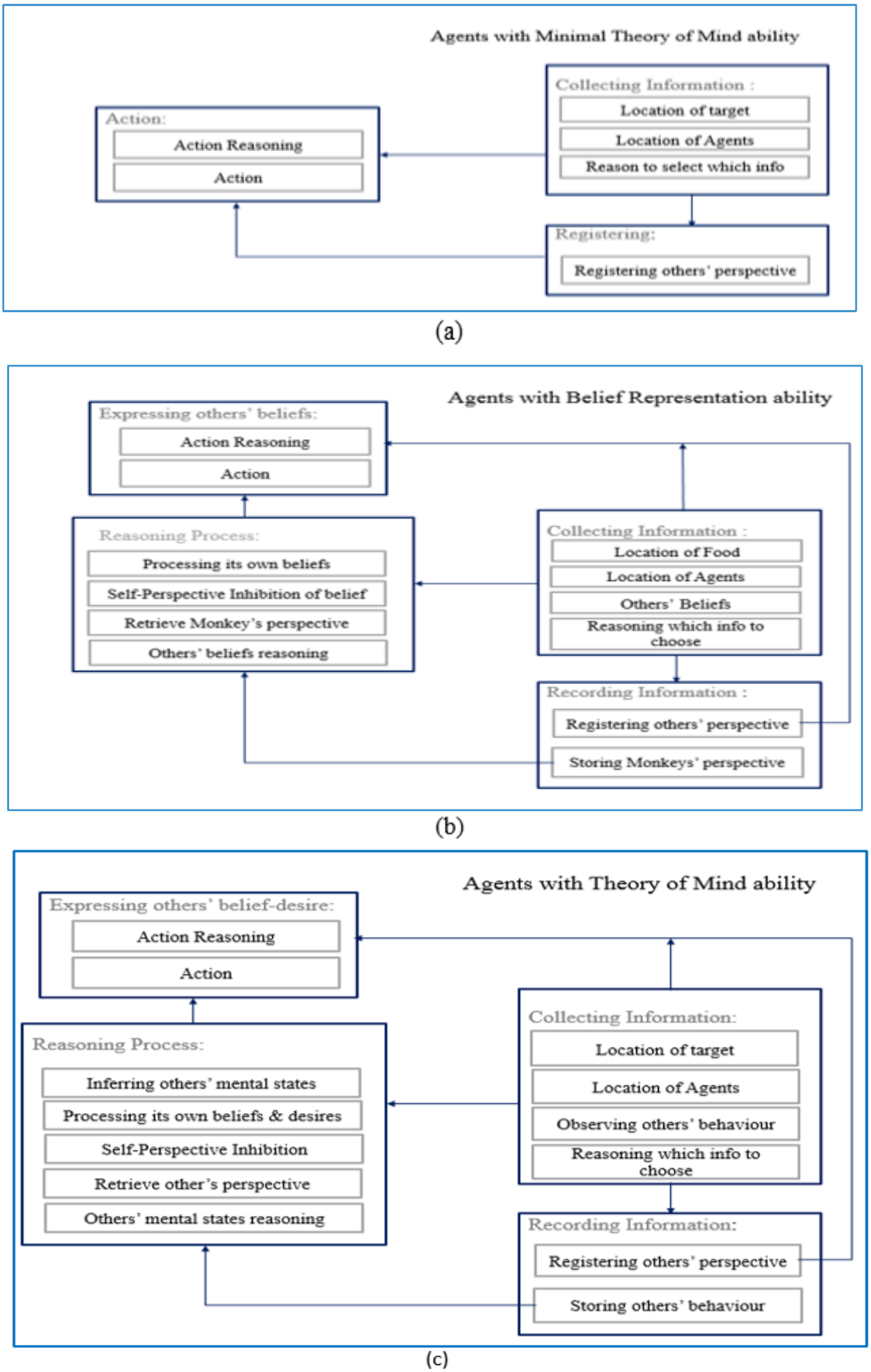


Figure 92. Agents' arrow and box diagrams with different levels of theory of mind ability

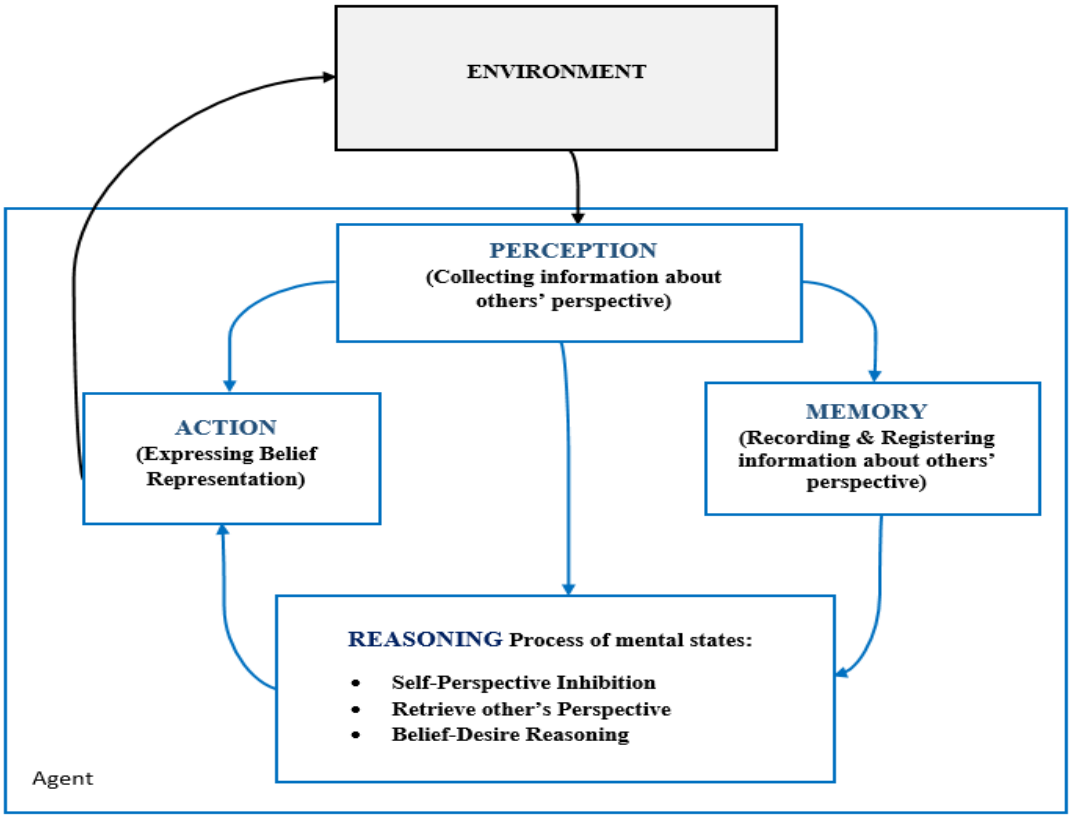


Figure 93. Belief Representation Systematic Approach (BRSA)

4.3 BRSA and its Applications

BRSA elaborates the underlying blocks and basic processes of belief presentation as precursors to a simple theory of mind, which is generalised into four phases:

- Perception (Collecting Information)
- Memory (Recording Information)
- Reasoning Process of Beliefs and Desires
- Action (Expressing others' mental states)

Three first steps including perception, memory and reasoning processes of beliefs and desires signify belief representation. The last additional phase, action, is a benchmark to express or test false belief understanding.

In addition, BRSA is beneficial as an analytical approach for distinguishing behavioral tasks from belief representation tasks, two characteristics that are not easy to separate.

BRSA is able to extract the beliefs and desires from other information through three subroutines in the reasoning phase. The behavioural tasks lack beliefs and desires to extract in the reasoning phase of BRSA. Thus, it is possible to distinguish the behavioural tasks from belief representational tasks.

Moreover, BRSA demonstrates the systematic link and the flow of information through its phases, as illustrated in Figure 93. The following sections of BRSA's applications provide some examples and more detail on this.

4.4 BRSA and Standard False Belief Task

Throughout the process of Sally and Ann false belief task, the child observes and collects the information from the world and updates this information to answer the questions such as:

- Where does Sally put the ball?
- Can Sally see the ball? (if Sally leaves the room)
- Where does Ann put the ball? (Sally is not in the room)
- Where will Sally search for the ball, when she returns?

BRSA offers a generic false belief algorithm, a structural approach to investigate false belief tasks core principles. This algorithm uncovers underlying phases by which one can successfully pass the false belief test.

The four phases of BRSA in Sally and Ann false belief task are as follows:

4.4.1 - Perception (collecting information)

The perception phase in Sally and Ann false belief task centres on the participant child who needs to collect supplementary information over time.

It is essential that the child correctly answer the above questions before any claim regarding the child's success in passing the false belief task. The participant child watches the Sally and Ann scenario, collects the information from her/his observation and updates its

information as the scenario continues. In fact, this critical phase is directly connected to all of other three phases and feeds them with online and updated information about the environment as illustrated in Figure 93.

4.4.2 - Memory (Recording information)

This phase, based on BRSA, in standard false belief task relates to storing information regarding others' perspective. The participant child is required to record Sally's perspective regarding the location of the ball (which is in the basket). Essentially, the child needs to store Sally's perspective information. Whilst Ann moves the ball to the box, the current location of the ball is no longer the same as Sally's perspective. Therefore, the child needs to allocate its memory to Sally's perspective information. Later in the task events, the child must retrieve this information to successfully pass the false belief test.

4.4.3 - Reasoning Process of Beliefs and Desires

Based on BRSA, this phase involves the information processing which includes:

1) Self-perspective inhibition; the participant child inhibits its own belief about the location of the ball.

2) The child must retrieve Sally's perspective from her/his memory.

3) The child reasons about Sally's desire; she wants the ball when she returns to the room. Therefore, the child is able to successfully pass the false belief test.

In summary, the participant child is supressing his/her own belief of the reality of the location of the ball while activating and retrieving Sally's perspective. A successful child who passes the false belief test, reasons that Sally's desire is to reach the ball and her belief regarding the location of the ball differs from the actual place of the ball.

4.4.4 - Action

The last phase of BRSA in standard false belief task is the question: "Where will Sally search for the ball when she returns?" Thus, the child should answer the question by

pointing at the object or answering verbally. This phase is referred as a measurement test (Bloom & German, 2000).

4.5 The evidence from theory of mind literature for BRSA

This section explores some of the evidence, gathered from theory of mind literature, which supports the BRSA phases. The evidence for each phase of BRSA is presented separately as follows:

4.5.1 - Perception (Collecting information)

The perception phase plays a crucial role in false belief tasks. Wellman (2014) explains that Sally has a false belief in the task because she did not see the exact key set of events; she did not see that the ball was moved by Ann. Therefore, Wellman (2014) includes "seeing leads to knowing" in false belief understanding. He uses the term "information access" to clarify how having access to information is critical and explains that pre-schoolers who pass standard false belief tasks are able to pass related seeing-knowing tasks. Similar to the perception phase of BRSA, the second principle of minimal theory of mind proposed by Butterfill and Apperly (2013) is concerned with the concept of field and encountering together as a substitute for perception phase. A field represents the agents' related area that includes objects. Whilst the object is in the agents' field of view, the relation of encountering occurs between agent and object, which have already described in general introduction.

In other words, agents are able to perceive other agents or objects in their field of view by encountering them; they collect the information in their field of view, which are critical for false belief task and for minimal theory of mind ability.

4.5.2 - Memory (Recording information)

Similar to BRSA, there is a strong body of evidence in child development research which suggests that at least a minimum level of working memory is necessary for false belief task

(e.g. Perner, 1991; Wimmer & Hartl, 1991; Davis & Pratt, 1995; Gordon & Olson, 1998; Keenan, Olson, & Zopito, 1998; Hughes, 1998; Keenan, 1998; Carlson, Moses, & Casey, 2002; Doherty, 2009; Slade & Ruffman, 2005; Apperly, 2012). The registered location of the object in protagonist's perspective in the standard false belief task is different from the real location of the object. To answer the false belief test question, retrieval of others' belief perspective from the memory might assist the belief representation process which have been stored through the task. (Hollebrandse, Hout, & Hendriks, 2014)

4.5.3 - Reasoning process of beliefs and desires

The reasoning phase of BRSA consists of three sub-routines: self-perspective inhibition, retrieving the data from memory and selective process of others' belief and desire. These sub-routines have been elaborated in the literature from different views. For example, false belief tasks encompass reasoning of everyday beliefs-desire psychology, a system of reasoning about mind, behaviour and their connection to the world that provides explanation and predictions of actions through the individuals' beliefs and desires (Wellman, 2014). Precisely Wellman describes that belief-desire psychology includes in one hand a range of concepts such as preferences that forms an individual's desires and in other hand, "perceptual-historical experiences" forms one's beliefs.

Moreover, the studies by Diamond (1991) and Carlson et al. (2002) have suggested that self-perspective inhibition together with working memory are associated with false belief task. It may be necessary to inhibit the reality information in a coherent belief representation (Apperly et al., 2007). The executive function is necessary for belief representation perhaps by inhibition of self-perspective (Leslie, Friedman, & German, 2004).

Research by Leslie et al. (2004) proposes theory of mind mechanisms (ToMM) consisting of three principles. The first principle relates to a meta-representational system for belief and desire as such, while the second principle explains a selection process (SP) system that

enables inhibition of own true beliefs. In the third principle, they devised the term "true-belief default" that explains the best guess of another's belief is that it is the same as one's own. Thus, to succeed in a false belief task, one must inhibit true-belief default. For example, to predict Sally's action, Leslie explains that the child must consider Sally's desire and beliefs. At first, the true-belief as default is the most salient. If the child is able to inhibit its own perspective, then the false belief becomes more salient and the child selects it (Leslie, Friedman, & German, 2004). Furthermore, there is a strong body of evidence that there is a correlation between executive functions (e.g. working memory, inhibitory control) and false belief task (Hughe, 2011a).

These arguments in literature are consistent with the BRSA's subroutines. Firstly, this phase of BRSA clearly illustrates Wellman's belief-desire psychology and the connection between beliefs and desires, which will lead to the actions. Secondly, self-perspective inhibition is a critical subroutine of this phase of BRSA which is compatible with the literature. Thirdly, the role of memory, retrieving the information from memory and using it in the process of theory of mind ability is another subroutine of the reasoning phase of BRSA.

4.5.4 - Actions

This phase results in an action based on understanding others' mental states such as their beliefs and desires. This action is based on three earlier phases in BRSA. This phase is parallel to the false belief task's test, and depending on the task, it comes in various forms. For example, in nonverbal false belief tasks eye gaze time or eye tracking can measure it. While in verbal false belief tasks, this is through a verbal response to a specific question such as "Where will Sally search for the ball?"

In theory of mind literature, this phase has been subject to considerable debate. For example, (Mitchell, 1996) argues that children acquire theory of mind before it can be

measured by a theory of mind task due to lack of sufficient executive function skills such as inhibitory control and memory. In fact, a measurement task for theory of mind is analogous to this phase of BRSA. The expressing phase might require more complex reasoning separate from understanding others' beliefs. This means, it is important to distinguish between theory of mind ability demands and theory of mind measurement task ones.

Hughes argues that the ability to understand or infer others' mental states differs significantly from the use of this information (Hughes, 2011-a). She states that recognition of others' mental states such as thought and feelings do not certainly certify concern for others. Besides, she emphasises an important distinction between having an ability and performing it in our behaviour. The study by Hughes argues that children with theory of mind ability use their understanding in different ways when interacting with peers. She claims that there are robust findings suggesting that the relationship between theory of mind ability and using this ability in social life is stronger for some children than for others (Hughes, 2011-a). The constructed concept of theory of mind is required in combination with the 'use' of the concept of theory of mind to show the link between theory of mind and peer relationships (Caputi et al., 2012).

In false belief task, Doherty (2009) clarifies that explanation is easier than prediction for children. The expressing phase includes different measurement tasks conveying different levels of complex utilities and reasoning. Hence, the complexity of the measurement task sometimes might be a barrier for the child to pass the false belief task.

Intriguingly, computational models and developmental literature consistently propose that mental states inferences and the actions influenced by these inferences are empowered by the principle of rational actions: "The expectation that agents act efficiently, within

situational constraints, to achieve their goals" (Jara-Ettinger et al., 2015, p. 1). Thus, the principle of rational action is applicable in the expressing phase.

In addition, another influential thread in false belief research proposes that there is a correlation between children's verbal ability and their false belief understanding (Astington & Jenkins, 1999; Watson, Painter, & Bornstein, 2001; Farrar & Maag, 2002). Nevertheless, Slade and Ruffma, (2005) suggest that early child language correlates with later false belief but not the reverse. This evidence is consistent with the BRSA argument that reasoning and expressing others' beliefs involves more than understanding it.

4.6 BRSA and the reasons for failure in false belief tasks

There are different reasons for failure in false belief tasks based on BRSA. These reasons include:

1. Lack of effective resources or apt reasoning in the expressing phase; for example, lack of apt rationality to achieve goals (e.g. in real life: deficiency in language development, linguistic ability and language comprehension).
2. Lack or impairment of rational reasoning in reasoning phase
3. Self –perspective inhibitory dysfunction
4. Memory problems in storing or retrieving data
5. Issues in perception phase such as incorrect information or attention problem.
6. A combination of these reasons

There is extensive literature on the role of memory in theory of mind (e.g. Fliss et al., 2015; Arslan et al., 2015; Hughes, 2011-b; Samson et al., 2004; Carlson et al., 2002). The self-perspective inhibition part in false belief task (e.g. Hughes, 2011-b; Leslie & Polizzi, 1998; Carlson, Moses, & Hix, 1998). The linguistic ability and language development role in false belief task (e.g. Schaafsma et al., 2015; Jones, Gutierrez, & Ludlow, 2015; De Villiers, 2005; Milligan, Astington, & Dack, 2007; Astington, 2001). Language

comprehension (Frank, Baron-Cohen, & Ganzel, 2015) and reasoning (Birch & Bloom, 2007) in support of the reasons of failure in false belief tasks in BRSA.

4.7 The link between minimal theory of mind and BRSA

The first principle of minimal theory of mind, goal directed action, is a pre-assumption in BRM and BRSA. This includes the general rule that each agent has a goal to move towards food. The second principle, field and encountering, is parallel to the perception phase of BRSA during which agents encounter food and other agents in its field of view. The registration and successful registration as the third principle of minimal theory of mind corresponds to the memory phase in BRSA. However, there is a subtle difference between registration and recording information, which relates to the length of time and the type of memory that the information stores. In registration, the information is only accessible for the current time step whereas in recording the information is accessible for future time steps. The last matching point is the action of the fourth principle in minimal theory of mind and expressing (action) phase in BRSA. The difference between these actions is that action in minimal theory of mind does not involve any reasoning whereas action in BRSA might involve with reasoning depending on the task.

Clearly, the reasoning belief and desire phase, which is the third phase in BRSA, does not exist in minimal theory of mind. The reason is that minimal theory of mind does not involve any reasoning and sophisticated concepts (Butterfill & Apperly, 2013). A comparison between BRSA phases and minimal theory of mind principles is illustrated in Figure 94.

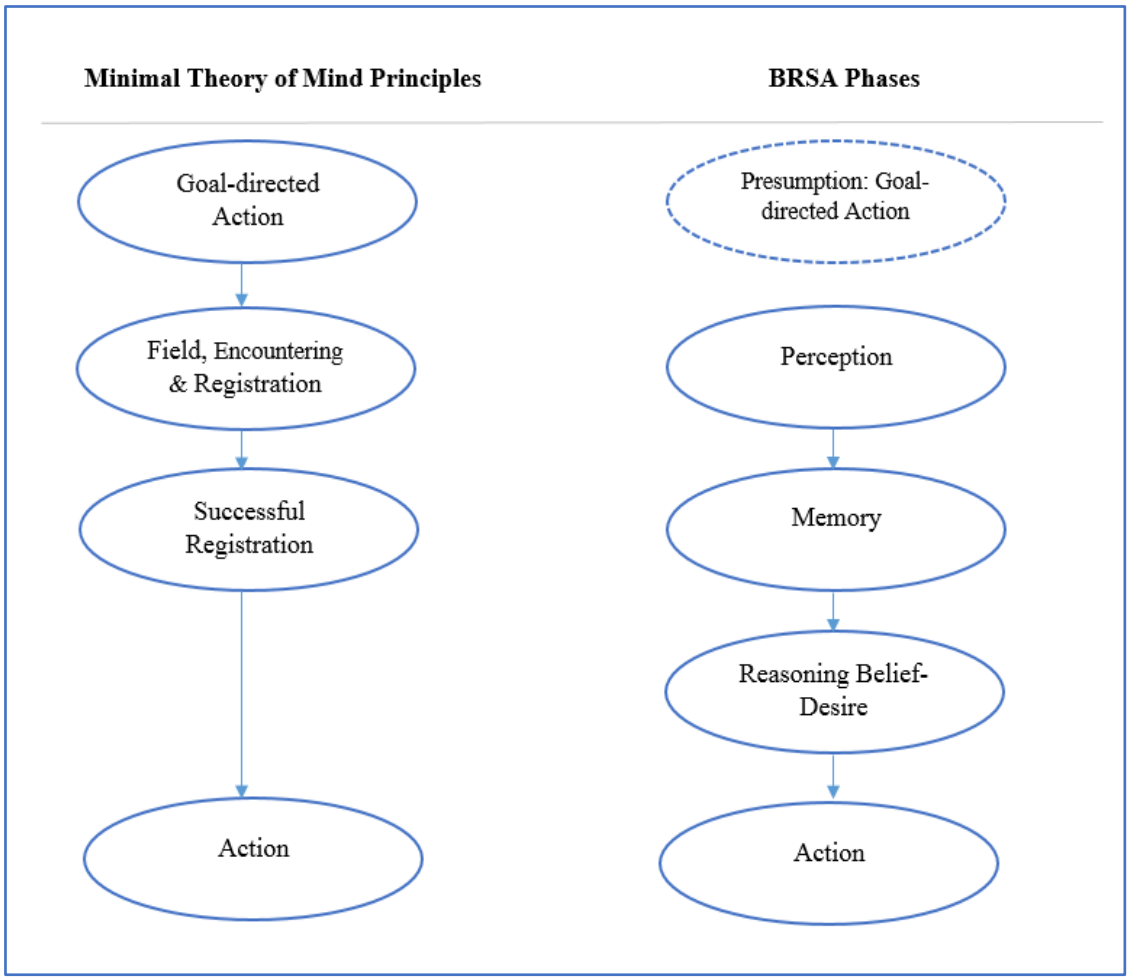


Figure 94. A comparison between minimal theory of mind principles and BRSA phases

4.8 In which conditions is false belief task a decisive test for theory of mind based on BRSA?

False belief tasks in the literature involves challenging actions, engaging with complex reasoning, intellectual connections or skills such as linguistic abilities, which might be more demanding than understanding others' beliefs. Particularly, children's false belief tasks are in this category.

As already alluded in section 2.5.3 and section 3.5.1, BRSA illustrates the demands on memory, inhibitory control and reasoning in its third phase. Moreover, BRSA shows that the expressing phase is a separate phase from understanding others' false beliefs. The

expressing phase which includes an action or a measurement test needs its own resources and separate reasoning, depending on the task.

The reason that the standard false belief task might not meet the conditions as an acid test for theory of mind is that firstly the reasoning phase of BRSA demands cognitive resources. The subroutines of reasoning, which include mental states reasoning, inhibition, retrieving the information about others' mental states perspective from memory, and the required interconnection between them are not basic resources. Therefore, having these resources is a pre-condition for success in false belief tasks.

Secondly, the expressing phase demands more than understanding others' beliefs.

Hence, BRSA validates the point that false belief tasks require sufficient resources as preconditions to act as a decisive test for theory of mind. Note that the concept of the theory of mind is different from minimal theory of mind as Apperly suggests. In other words, if the precondition of resources is met, then a false belief task is a decisive test for higher level than a minimal theory of mind.

4.9 In which conditions is false belief task a decisive test for theory of mind in the literature?

The false belief task is used extensively and recognised by researchers as an acid test for theory of mind ability (e.g. Wellman & Bartsch, 1988; Doherty, 2009; Apperly, 2011; Workman & Reader, 2014). The central aspect of false belief task is that the child needs to predict others' behaviour by inferring their perspective, which is different from the current world. For example, to successfully pass the standard false belief task, the child must focus on Sally's perspective rather than the present location of the ball. This is the underlying reason that understanding others' false belief has been considered as an acid test for the presence of theory of mind ability.

On the contrary, Bloom and German (2000) advocate the abandonment of the false belief as an acid test for theory of mind. Their critical paper explains two reasons that false belief task needs to be abandoned as a test of theory of mind. Firstly, they propose that to pass the false belief task requires abilities beyond those of theory of mind. They point out the difficulties of the false belief task for example; the child needs to follow the actions of Sally and Ann such that remember the locations of the ball which Sally and Ann used. In addition, the child needs to comprehend that Sally could not have seen the ball's new location as she left the room. Furthermore, the child has to fully understand the test question. Bloom and German also state that the false belief task is impossible to understand for a child of 2 years of age and hard for 3 year old children due to insufficient attentional and linguistic resources. Moreover, Bloom and German argue that it is important to simplify the task or question as is shown by the results from some studies (e.g. German & Leslie, 2000; Surian & Leslie, 1999; Carlson, Moses, & Hix, 1998; Freeman & Lacohee, 1995; Moses, 1993; Mitchell & Lacohee, 1991; Freeman, Lewis, & Doherty, 1991; Lewis & Osborne, 1990). These studies show that children might often succeed at the age of 3 which indicates that developmental change happens earlier than expected. However, as already indicated Wellman's meta-analysis (2014) show no consistent effect. Secondly, Bloom and German analyse that an autistic individual fails the false belief task due to the lack of theory of mind, whereas a typical 3 year old child who has sophisticated ability to reason about mental states might fail the task because of inefficient processing abilities (Bloom & German, 2000).

Conceivably, the subject child in false belief task needs to first understand the measurement test question. The link between the false belief task scenario and the measurement test question requires a certain 'level of reasoning'. The reasoning level is critical for false belief task to be decisive for theory of mind. When the reasoning level and cognitive

resources such as memory is higher than the subject child ability in the false belief task, the condition that false belief task is a decisive test for theory of mind is not complete.

4.10 Minimal theory of mind as a decisive construct for theory of mind based on BRSA

One important advantage of minimal theory of mind construction includes its capability to demonstrate systematic success on measurement tests for theory of mind including many false belief tasks (Butterfill & Apperly, 2013). On one hand, minimal theory of mind concerns someone "with limited cognitive resources or little conceptual sophistication" (Butterfill & Apperly, 2013, p. 1) and on other hand, "Where a task goes beyond these limits, we can be sure an agent is not using minimal theory of mind only." (Butterfill & Apperly, 2013, p. 6). Minimal theory of mind does not include reasoning phase of BRSA. More precisely, the first three principles of minimal theory of mind are critical for understanding beliefs-like and are not for the purpose of representing psychological attitude as such. The fourth principle is concerned with an action that expresses the understanding of others' simple beliefs. The action in the fourth principle does not involve any complex resources and it is free from any sophisticated demands. On this basis, BRSA and its analysis of phase three and four, concerning reasoning and an action respectively, as described earlier in this chapter, supports Apperly and Butterfill's claim that minimal theory of mind is capable of explaining systematic success on tasks that are supposed to be decisive tests for theory of mind.

CHAPTER 5

5. GENERAL DISCUSSION

5.1 GENERAL DISCUSSION

A meaningful social life relies on understanding others' mind and behaviour as well as our own. Theory of mind is the ability to reason about an individual's mental states such as beliefs, desires and intentions, and to understand and predict how these mental states shape an individual's behaviour.

This dissertation has offered a computational framework to develop a systematic understanding of simple theory of mind competence. Agent-based models (ABMs) provide a computational platform to simulate individual or collective autonomous entities, as agents and their actions. Agent-based simulations evaluate the effect of agents' interactions within their environment. ABMs are robust technical laboratories for assessing agents on micro and macro levels. For this purpose, Repast Simphony, a Java based simulation platform, was used to design and implement two agent-based models; one to simulate belief representation based on standard false belief tasks (BRM) and the other to simulate a simple theory of mind (MSM). At first glance, an agent-based model for theory of mind seems eccentric in developmental literature as theory of mind research methodology traditionally relies mostly on experimental and brain imaging studies. However, cognitive science can greatly benefit from agent-based models for analysing cognitive processes and social aspects of cognition through agents' interactions (Sun, 2006). Throughout this study, the analyses of BRM and MSM were originally dedicated to agents' performance in a virtual competitive society, and more importantly, the core procedures of understanding others' mental states. An agent's interactions in the environment manifest as two types of actions; their own movement and other agents' movement, which influence each other. The first model, BRM, consists of three different types of agents that move in their neighbourhoods to consume food: Infant agents, Monkey agents and Control agents. While Infant agents are able to understand Monkey agents' perspective regarding the location of food and store

this information, Monkey agents are only capable of storing the location of food. Control agents are used as a control measurement and they are able to track Monkey agents' fields of view. The performances of agents are assessed based on their ability to consume food. The results of the simulations show that Infant agents perform higher than Monkey agents do. In addition, Monkey agents' performance is higher than Control agents, particularly in situations in which agents use their strategies often.

The second model, MSM, consists of six types of agents with two mutually exclusive mental states regarding the target: Active or Passive. Agents, including Random, Food, Control, MinToM, Infer and Reality agents, move to achieve a target when they are in an Active state, whereas they cannot achieve any target while in a Passive state. Infer agents are able to infer other agents' mental states from their behaviour in achieving a target and storing this information. MinToM agents register other agents' perspectives regarding the location of food. Control and Food agents are both able to only sense the location of food, but they differ in terms of their expansion of field of view and their strategies when they are in a Passive state. Random agents move randomly, and Reality agents' strategy is similar to Infer agents but they do not infer others' mental state, instead they have direct access to them. Agents' performances are evaluated based on the number of time steps that agents are in an Active state but are not successful in achieving a target. The results of simulations show the descending order of performance is Reality, Infer, MinToM, Control, Food and Random agents.

The case in point is that agents with theory of mind ability (Infer agents) in MSM decide their own movements while simultaneously interpreting others' movements, which are associated with both an agent's own mental state (such as beliefs and desires) as well as others' mental states. In the same vein, Infant agents in BRM are capable of representing

their rival agents' (Monkey agents) beliefs. Monkey agents are unable to recognise others' beliefs.

Thus, both models' simulation results have consistently suggested that understanding others' mental states has an impact on successfully achieving goals in a competitive society. In general, the results show that agents with theory of mind ability perform better than agents that have minimal level, which in turn perform better than those agents that lack this ability. However, there are uncertain environments in which agents are unable to use their abilities and apply their strategies leading to different results. For agents with theory of mind ability to perform effectively, a rational link between the belief representation and the related action (the action that expresses the understanding of others' belief) is required. This reveals that the 'use' of understanding others' mental states requires rational reasoning.

Whilst the results of MSM demonstrate higher efficiency for agents with theory of mind ability, the agents capable of tracking others' field of view (MinToM) are the second highest in performance. These results suggest that considering others' field of view is the minimum scaffolding needed to understand others, enabling agents to track others' belief-likes which is consistent with Butterfill & Apperly's (2013) minimal theory of mind.

There is a synchrony between the information that each agent perceives then uses in processing inferences of others' mental states and the level of that agents' theory of mind competence. As the complexity of perceived information and the processes become more advanced, agents' performance increases: initially from understanding their own mental states, then tracking others' field of view and finally making inferences about others' mental states. There is a coherent association between the agents' level of theory of mind, their strategy and the necessary information they require.

The largest improvement in agent performance (excluding Random agents) occurs from MinToM agents to Infer agents, signifying the importance of understanding others' mental states in a virtual society. Due to bounded rationality features and uncertain environments in MSM and BRM, even Reality agents with direct access to mental states do not perform perfectly.

5.1.1 BRSA in a nutshell

This thesis presents a set of basic processes for a simple theory of mind, derived from BRM and consistent with MSM. The implementation of these two models has revealed the foundation units of theory of mind processes which is called Belief Representation Systematic Approach (BRSA).

BRSA illustrates the underlying processes that are identical across a variety of theory of mind tasks including false belief tasks. BRSA is powerful enough to address some of the ambiguity behind theory of mind tasks by breaking a task into the standard phases that clarify its processes. BRSA is also able to simplify and explain some of the complex characteristics within the tasks. It unifies various credible findings in the literature.

BRSA represents the main basic blocks of a simple theory of mind procedure within four phases: perception, memory, reasoning beliefs and desires, and action.

The first phase involves agents' perception (collecting information from the environment). This mainly corresponds to information about others' mental states as well as the location of targets. This phase is consistent with the second principle of minimal theory of mind, which regards field of view and encountering, (Butterfill & Apperly, 2013) that enables agents to see other agents or objects in their field of view by encountering them. Agents collect information, and then recognise the link between the object and the information in this phase. Perception phase includes reasoning to determine what information regarding

the target is more important to select and in what priority. For example, the information about the nearest targets and the agents with a shared target has the highest priority.

Similarly, one of the social information gathering skills in human infants in real life is decoding the social environment information and discerning the information about an entity such as an object (Baldwin & Moses, 1996) which is analogous to this phase. In addition, information access is a precondition for pre-schoolers to be able to pass false belief tasks (Wellman, 2014) by perceiving their environment. This is analogous with the first phase of BRSA in BRM, in which Infant agents need to gather information about Monkey agents' beliefs regarding the location of the food.

Accordingly, Infer agents collect information by observing other agents' actions when reaching a target in MSM. In general, the perception phase is interconnected with other phases of BRSA as an online access to the information.

The second phase of BRSA involves storing the information into agents' memory and using it in future time steps (for the reasoning of others' mental states). Thus, memory is indispensable in this phase of BRSA, which is analogous to the need for memory in false belief tasks (Apperly et al., 2007) and memory is one of multiple domains in social understanding processes (Mitchell, Macrae, & Banaji, 2004). For example, the key point in BRM relates to the location of the food; the real location of food is different from the protagonist agent's (Monkey agents) perspective, which has already been stored in the agent's (Infant agents) memory.

Likewise, MSM elucidates memory demands on Infer agents to store necessary information. This information will be used to infer other agents' desires and beliefs.

Both models demonstrate the importance of memory resources in the process of belief representation.

The third phase of BRSA is called the reasoning process of beliefs and desires, and encompasses three subroutines: the selective process of beliefs and desires, inhibition and retrieving data from the memory.

In the first subroutine, agents start to reason about their own beliefs and desires. For example, in BRM, Infant agents' desire is achieving food, whereas their beliefs about the location of food changes due to the dynamics of the environment. In fact, the present information creates their new beliefs, and their information about others' beliefs will be stored into their memory and will be used in the next time step. These "perceptual-historical experiences" (Wellman, 2014, p. 24) form the Infant agents' beliefs about others' perspectives. Subsequently, Infant agents need to reason about others' beliefs and desires: Monkey agents' desire is to achieve food. Monkey agents' beliefs, when they no longer see the previous food, remain the same (for one time step). This might result in a contradiction between the real location of the food and the Monkey agent's belief. Therefore, in this situation, there are two different beliefs about the location of one food, Infant agent's belief and that of the Monkey agent. Thus, Infant agent needs to temporarily inhibit its own true belief about the location of food, and consider Monkey agent's belief from its memory (self-inhibition subroutine). Intriguingly, the role of inhibition is prominent in false belief situations.

The last subroutine relates to retrieving information from the protagonist's perspective, which was stored in the memory in the second phase. For example, Infant agent retrieves Monkey agents' perspectives regarding the location of the food.

One significant difference between the two models appears in the reasoning phase: MSM uses a higher level of reasoning than BRM. This is because, in MSM, agents infer others' mental states by observing their actions, whereas, in BRM, agents directly observe others' beliefs with no inferences. This suggests that the reasoning phase determines the level of

complexity of the theory of mind task. However, the subroutines of the reasoning phase are indispensable for false belief and theory of mind tasks and new subroutines can only be added to the reasoning phase as its complexity increases.

Undoubtedly, the reasoning phase indicates a central information-processing step for theory of mind competence. In addition, the contribution of the executive function to belief representation (Apperly, 2011) has been elucidated by the reasoning phase of BRSA:

1) The role of 'memory' in retrieving the information subroutine,

2) The role of 'inhibitory control' in self-perspective inhibition subroutine,

3) The role of 'reasoning' in selective process of belief and desire subroutine

Indeed, belief representation is accomplished by the end of the reasoning phase. However, agents have not exhibited any evidence of this understanding in their behaviour yet.

The last phase of BRSA relates to expressing the understanding of others' mental states as an action or behaviour. In other words, performing an action by using a mental representation (Hughes, 2011-a). This phase is analogous with the measurement test (for example, a question like 'Where will Sally look for the ball?' in the Sally and Ann false belief task) in false belief tasks.

In real life, the action of this phase might be as simple as an eye gaze and eye tracking or a complicated action that requires separate reasoning from the reasoning phase. Therefore, the distinction between having theory of mind ability and using this ability in an action is important to prevent the existing confusion about theory of mind measurement tests, such as the linguistic ability necessary for the false belief task in children.

In general, it is possible to express understanding of others' mental states with different actions, as has been shown for Infer agents in MSM. 'How to use' this ability might be an obstacle to pass the measurement test of theory of mind. Besides, by considering 'how to use' this ability, the principles of rational actions is essential.

BRSA elucidates the delicate boundary between understanding mental representation and expressing or using this representation that might sometimes cause confusion in the literature. It is valuable to articulate one of these common issues concerning false belief task below.

5.1.2 False belief task as a decisive test for theory of mind

In spite of the fact that false belief tasks have been widely used as a decisive test for theory of mind ability, the contrary discourse argues that success on false belief tasks might need more than understanding others' mental states (Bloom & German, 2000; Baron-Cohen, Leslie, & Frith, 1985; Hughes, 2011-a). This view is compatible with BRSA, which demonstrates: firstly, the reasoning process of beliefs and desires phase is demanding in cognitive resources such as memory and inhibitory control, and secondly, expressing others' belief (action phase), which is equivalent with the passing false belief tests, might involve more than understanding others' belief. Thus, the third and fourth phase of BRSA determine the level of reasoning and the necessary resources for false belief task preconditions. Therefore, for a false belief task to be an acid test, it requires preconditions of resources and reasoning. This demonstrates an important condition on interpreting the relationship between patterns of success or failure on false belief tasks in terms of theory of mind abilities.

Similarly, to pass false belief tasks in real life, children need to link complex series of events (Moses, 2001). Therefore, children may fail to pass false belief test due to requiring higher order control to express their belief representation competence (Moses, 2001). This corresponds with the findings about the level of reasoning and resources required in reasoning phase and expressing phase of BRSA.

5.1.3 Reasoning role in theory of mind

In line with the conclusion from BRM and MSM, the role of reasoning is highlighted in the third and fourth phases of BRSA for theory of mind and false belief task. This shows that improvement in the complexity of reasoning and cognitive resources reinforces the improvement in both forms of understanding others' mind and using this understanding. This also suggests that, as the complexity and amount of information regarding others' mental state increases, the level of reasoning can also increase.

In general, BRSA in a simple theory of mind model (MSM) replicates the same procedure in Belief Representation Model (BRM), applying similar phases except with more reasoning for inferring others' mental states.

One result of BRSA and its analysis lends weight to the reasoning phase in theory of mind processes and shows its heuristic nature. However, the link between theory of mind and reasoning still is an open question in the literature. For example, research by Vaart and Hemelrijk (2014) questioned how sophisticated reasoning level in animals is, regardless of whether they have a theory of mind.

5.1.4 BRSA and minimal theory of mind

This study also seeks to provide clarity on how BRSA, which represents basic theory of mind procedure through its phases, particularly the reasoning phase, responds to minimal theory of mind principles; the ability to track others' perception, knowledge and belief-likes with limited cognitive demands and without conceptual sophistication (Butterfill & Apperly, 2013). In general, minimal theory of mind is constructed upon four principles: goal directed action, field and encountering, successful registration and action influenced by the registration. For example, in BRM, Control agents and Infant agents satisfy these four principles: the agents' goal is to consume food (Principle 1), the agent has a field of view to observe and encounter objects in its limited neighbourhood (Principle 2), agents

register the location of the food in their field of view as a belief-like (Principle 3), and agents track other agents' beliefs about the location of food in their field of view and use it to take an action (Principle 4). Based on these principles, Control agents and Infant agents possess minimal theory of mind ability. However, Infant agents are using more than minimal theory of mind. This suggests that minimal theory of mind is the minimal sub set of understanding others' beliefs. Similarly, it suggests that MinToM agents in MSM have minimal theory of mind ability.

Furthermore, a scrutiny on BRSA phases reveals the link and difference between belief representation process and minimal theory of mind principles. Firstly, goal directed action is a presumption in BRSA, which is associated with the first principle of minimal theory of mind. The second principle relates to field and encountering, which corresponds to perception phase of BRSA. The third principle of minimal theory of mind, registration and successful registration, is parallel to the memory phase of BRSA. There is a delicate difference between these two concepts; registration in minimal theory of mind refers to registration of other's perspective and use of it in the current time step, whereas memory phase in BRSA reflects both recording others' perspective in agents' memory for using in the next time steps, and registration. This difference shows the role of mental time traveling (using information in the current and future time steps) and different memory demand (short-term and long-term) in theory of mind. The fourth principle in minimal theory of mind regards an action that is influenced by registration, which is compatible with the action phase in BRSA.

Since minimal theory of mind does not involve mental representations as such and sophisticated concepts (Butterfill & Apperly, 2013), one difference between BRSA and minimal theory of mind is that phase four of BRSA and its subroutines, involving reasoning

process of beliefs and desires such as self-perspective inhibition, do not exist in minimal theory of mind.

Apperly and Butterfill argue that someone with minimal theory of mind ability could pass the tests, which are supposed to be a decisive test of theory of mind ability including false belief tasks. Through the lens of BRSA and its analysis, their argument is strong and credible for two reasons; firstly minimal theory of mind concerns someone "with limited cognitive resources or little conceptual sophistication" (Butterfill & Apperly, 2013, p. 1) and "Where a task goes beyond these limits, we can be sure an agent is not using minimal theory of mind only." (Butterfill & Apperly, 2013, p. 6). This again highlights that minimal theory of mind does not include phase three of BRSA and the resources and the preconditions that it requires. Secondly, the fourth principle, which is parallel with phase four of BRSA, includes an action without involving complex resources or reasoning levels. Therefore, the principles of minimal theory of mind exclude the two preconditions of resources that are essential for false belief task to be an acid test.

This concludes that BRSA covers both of the two systems of humans' belief tracking (Apperly & Butterfill, 2009) by removing the central process of reasoning phase and taking into account the time and memory differences in registration principle in minimal theory of mind and phase two of BRSA.

5.1.5 BRSA as a filter for behavioural tasks

This thesis concludes that BRSA is capable of identifying pure belief representational tasks from behavioural tasks, which would otherwise cause ambiguity in experiments. The reason for this capability is that BRSA phases give structure to the task, and are able to logically extract one's beliefs and desires from other information. The analyses of beliefs and desires in the reasoning phase and its three subroutines provide the extraction steps for beliefs and desires. In this way, the behavioural tasks lack the analyses of beliefs and

desires and reasoning phase in BRSA. Therefore, the BRSA phases filter the behavioural tasks from belief representational tasks.

5.1.6 Theory of mind for planning

Another aspect of simulation results of MSM demonstrates that by reducing the reasoning and planning beyond the action of Infant agents, the efficiency of Infant agents' performances decreases sharply. Essentially, theory of mind information feeds actions with planning for future time steps, resulting in higher performances in agents with theory of mind capability. This crucial distinction between theory of mind competence and the use of it in actions has already been suggested in a children's study by Hughes (2011-a). Moreover, children use their theory of mind understanding in different ways in their social life with their peers (Hughes, 2011-a) which relates to how we 'use' the information about understanding others' mental states in the action phase of BRSA.

5.1.7 Imperfect perception

Intriguingly, there is a general rule in perception in both models, BRM and MSM, which provides an environment that agents' access to information is not complete and perfect. Observed information is more precise the closer it is to the agent, and so the information perceived at the edge of an agent's field of view is incomplete in comparison the information perceived near the centre of an agent's field of view. Although both pieces of information are in the field of view, their accuracy is not at an equal level. This is because the amount of information (about other agents) which is outside an agent's field of view increases as the distance from its centre of field of view increases. The agents' imperfect vision has a direct impact on their perception of the environment.

5.1.8 Resources and costs

BRSA demonstrates that perception, memory, inhibition, and selective process reasoning form the main components of belief representation processes. Besides, it also shows that

expressing others' beliefs might require reasoning. Collectively, these components form a network of resources (Schaafsma et al., 2015; Mohnke et al., 2015; Gallagher & Frith, 2003; Carrington & Bailey, 2009) to represent others' beliefs.

Furthermore, understanding others' mental states in MSM comes at the cost of time and resources. The results of MSM demonstrate that Infer agents are slower since they require more time to reason and retrieve information about others' mental states. Similarly, BRSA shows the inferring process requires interconnected resources of perception, memory, inhibitory control and reasoning.

5.1.9 Uncertainty

The interaction between agents and the environment enhances the dynamics of the world and has a direct impact on agents' beliefs regarding the location of food and also their decisions and actions. Together, this dynamic nature of the environment and its randomness (such as random placement of the food) escalates uncertainty and complexity in the environment. The other factors that determine the uncertainty of the environment are the number of agents, the number of targets and the ratio between them. If the uncertainty in the environment is large, agents may be unable to use their strategies.

5.1.10 Summary

In conclusion, in this thesis, the key concepts of understanding others' beliefs and a simple theory of mind are simulated in two different dynamic environments, in BRM and MSM, respectively.

The results of two models indicate that understanding others' mental states such as beliefs and desires have a direct positive impact on agents' performance in a competitive society. The link between "micro motivates and macro behaviour" (Schelling, 2006) in both models demonstrate a pattern from the individual agents' rules and abilities (in terms of theory of mind) to the efficiency of their performances at macro level.

The overall basic blocks of belief attribution processes and also a simple theory of mind process is identified and is depicted in a Belief Representation Systematic Approach (BRSA). The underlying building blocks and fundamental processes of theory of mind (Schaafsma et al., 2015) are elaborated by BRSA into four phases. The three first phases (perception, memory, reasoning of beliefs and desires) signify belief representation. In addition, the action phase is a benchmark to measure or use the understanding of others' mental states.

5.2 Limitations and future work

This thesis proposes two innovative agent-based models that demonstrate that agents with higher level of theory of mind ability perform better in achieving their goals in a competitive society. More importantly, these computational models offer a novel belief representation systematic approach for the underlying processes of theory of mind ability and false belief tasks. The shared sets of basic processes of different theory of mind tasks (Schaafsma et al., 2015) are defined through four phases in BRSA. This is a step forward towards clarifying the basic structure of theory of mind ability, which is often inconsistent in the literature.

This study is stimulating because it attempts to understand two distinct dimensions of the demands of mental states; in BRM, agents focus on others' beliefs because their desires are the same, whereas, in MSM, agents infer others' desires first. Although this might be considered as a drawback in MSM, it demonstrates that desires might have priority over beliefs and this is not truly a limitation.

Agents in MSM have either an Active or a Passive mental state in each time step. In other words, the possibility of different degrees between Active and Passive is not considered. At first glance, this approach of two mental states appears to be very simple. However, the simple modelling of the two states leads to a deeper practical understanding than if an

analogue model was used, as the analogue theories that claim to be closer to reality become more complicated in the end (Minsky, 1988). However, it would be reasonable to extend the model with degrees between the two mental states and examine the results by applying complexity theory.

Whilst the models work for a reasonable number of agents, when attempting to increase this to a very large number of agents, for example one million agents, it becomes computationally intensive to calculate all of the interactions and actions of agents and reduces the simulation running speed to a halt. Undoubtedly, there are computational ways to tackle this problem, but this does not affect the current results, which are sufficient.

In addition, this study highlights that reasoning plays a key role in theory of mind and individual differences in expressing understanding of others' mental states. However, more work is needed to establish when, why, and how reasoning methods are different in theory of mind tasks in individuals considering the principle of rational action.

BRSA is a generic framework for belief representation. Thus, it is feasible to utilize BRSA in many applications of ToM ability. For example, the game and gamification design involving understanding others' beliefs and desires would benefit from applying the BRSA structure to the players' performances.

Both models presented in this thesis have applied discrete time design fulfilling the aim of this study. Nevertheless, continuous time could be explored in future work. In fact, by applying continuous time, it is possible to measure the time consumed by each type of agent to process others' beliefs and desires. Thus, one option for future work regarding continuous time, is to examine how fast agents with lower levels of theory of mind act as opposed to agents with higher levels of theory of mind. Based on BRSA, agents with higher levels of theory of mind require more time to collect and store information about others' mental states and to reason than agents with lower levels. For example, agents with only

minimal theory of mind ability might be faster than agents which are able to infer others' mental states.

Past theory of mind research has been heavily experimental, rather than using computational models. Therefore, a useful direction for future work could be to perform more fine-grained analysis by undertaking a similar experiment with human participants and a comparison between these results and the virtual ones.

This study outlines simulation results and mainly focuses on the processes of theory of mind ability. There are, nevertheless, other approaches which are beneficial to our understanding of underlying theory of mind cognitive processes; For example, Bayesian theory of mind, Markov decision processes and probabilistic inferences about partially observable events, which consider causal inferences regarding belief, desire and action.

Besides, another important approach would be to implement agents which can learn to improve their understanding of others' mental states in both cooperative and competitive environments, and examine the effect of learning in theory of mind from a variety of angles.

The models proposed in this study, particularly BRSA, could not be explained under a specific account of theory of mind such as theory-theory. Indeed, this study was not intended to prove one specific account of theory of mind, and instead has demonstrated that coherent aspects from different accounts of theory of mind could be integrated together. Therefore, more research is needed to focus on details of different characteristics in each account.

BRSA provides the basic cognitive processes of theory of mind. However, it is possible to apply a structural cognitive architecture such as the CLARION to advance our understanding of social cognitive features of theory of mind. CLARION is an integrative architecture consisting of functional subsystems (e.g. the action-centred subsystem, the metacognitive subsystem, and motivational subsystem) with two dual implicit and explicit

representational structures for each subsystem (Sun, 2006). Nevertheless, CLARION is not completely operational yet. This indicates that agent-based modelling in psychology is still in its first steps, and by utilizing AI and complexity theory, it will improve substantially.

5.3 Conclusion

This thesis has implemented two computational models for false belief tasks and theory of mind ability. Firstly, this thesis offers a novel standardisation for the building blocks of theory of mind processes, a Belief Representation Systematic Approach (BRSA). This is the first attempt in the field, at least in the author's knowledge, in which the process of belief representation has been organised in a collective systematic set of phases, a methodological framework for a variety of theory of mind tasks including false belief tasks. It consists of four indispensable and linked phases for processing others' mental states; collecting information as perception, recording information in memory, reasoning process of belief and desires, and finally, expressing others' mental state as an action. Thus, BRSA phases identify a network of resources for theory of mind competence including perception, memory, inhibitory control and reasoning resources. Secondly, the models demonstrate how theory of mind ability improves the agents' performance in achieving their goals in a virtual competitive environment compared to agents with minimal theory of mind, which in turn perform better than those agents that lack this ability.

Intriguingly, minimal theory of mind principles for someone with limited cognitive resources (Butterfill & Apperly, 2013), are a sub set of BRSA phases. To understand the link between BRSA and minimal theory of mind, minimal theory of mind needs to remove the reasoning phase in BRSA and consider the time difference between registration and recording information in memory phase. This indicates that two systems proposed for human belief tracking (Apperly & Butterfill, 2009) are applicable in BRSA as a generic approach.

This study concludes that BRSA is able to distinguish pure belief representational tasks from other tasks such as behavioural tasks, which are problematic. BRSA places emphasis on the reasoning phase in theory of mind processes and shows its heuristic nature.

The author's expectation, therefore, is that BRSA, a novel standard approach to analysis of theory of mind processes, will be beneficial in future research.

REFERENCES

Abrahamson, D., & Wilensky, U. (2005). Piaget? Vygotsky? I'm game! Agent-based modeling for psychology research. *Paper presented at the annual meeting of the Jean Piaget Society*. Vancouver, Canada.

Airenti, G. (2015). Theory of mind: a new perspective on the puzzle of belief ascription. *Frontiers in Psychology*, *6*.

Apperly, I. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Hove: Psychology Press.

Apperly, I. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology*, *65*, 825–839.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-Like states? *Psychological Review*, *116*(4), 953–970.

Apperly, I. A., Back, E., Samson, D., & France, L. (2007). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *103* (2), 300-321.

Apperly, I. A., Samson, D., Chiavarino, C., Bickerton, W.-L., & Humphreys, G. W. (2007). Testing the domain-specificity of a theory of mind deficit in brain-injured patients: evidence for consistent performance on non-verbal, "realityunknown" false belief and false photograph tasks. *Cognition, 103* (2), 300-321.

Arslan, B., Wierda, S., Taatgen, N., & Verbrugge, R. (2015). The Role of Simple and Complex Working Memory Strategies in the Development of First-order False Belief Reasoning: A Computational Model of Transfer of Skills. *Proceedings of the 13th International Conference on Cognitive Modeling*, (pp. 100-105).

Astington, J. W. (2001). The Future of Theory-of-Mind Research: Understanding Motivational States, the Role of Language, and Real-World Consequences. *Child Development, 72*(3), 685–687.

Astington, J., & Jenkins, J. (1999). A longitudinal study of the relation between language and theory of mind development. *Developmental Psychology, 35*, 1311–1320.

Baillargeon, R., Scott, R. M., & He, Z. (2010). Trends in Cognitive Sciences. *False-belief understanding in infants, 14*, 110–118.

Baker, C. L. (2012). Bayesian Theory of Mind : modeling human reasoning about beliefs, desires, goals, and social relations. *DSpace@MIT*.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*.

Baldwin, D. A., & Moses, L. J. (1996). The Ontogeny of Social Information Gathering. *Child Development, 67*(5), 1915-1939.

Baron-Cohen, S. (1989). Are autistic children behaviourists? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders* .

Baron-Cohen, S., Jollife, T. M., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*, 812-822.

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

Bersini, H. (2012). UML for ABM. *Journal of Artificial Societies and Social Simulation , 15 (1)* (9).

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, , 18*(5), 382–386.

Bloom, P., & German, T. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25-31.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, 2*(1), 14-23.

Brousmichea, K. L., Kanta, J. D., Sabouretb, N., & Prenot-Guinardc, F. (2016). From Beliefs to Attitudes: Polias, a Model of Attitude Dynamics Based on Cognitive Modeling and Field Data. *Journal of Artificial Societies and Social Simulation, 19*(4).

Butterfill, S., & Apperly, I. (2013). How to construct a minimal theory of mind? *Mind and Language*, *28*(2), 606-637.

Call, J., & Tomasello, M. (1999). A nonverbal theory of mind test. The performance of children and apes. *Child Development, 70*, 381–395.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12* (5), 187-192.

Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal Effects of Theory of Mind on Later Peer Relations: The Role of Prosocial Behavior. *Developmental Psychology, 48*(1), 257-270.

Carlson, S. M., Moses, L. J., & Casey, B. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*(2), 73–92.

Carlson, S., Moses, L., & Hix, H. (1998). The role of inhibitory control in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping, 30(8)*, 2313-2335.

Casti, J. L. (1997). *Would-Be Worlds: How Simulation is Changing the Frontiers of Science*. New York: Wiley.

Chin, K., Gan, K., Alfred, R., & Anthony, P. &. (2014). Agent Architecture: An Overview. *Transitions on sciences and Technology, 1*(1), 18-35.

Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. the western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B, 362*, 507–552.

Csibra, G., Bıró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27* , 111–133.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*(1), 25-31.

De Villiers, J. G. (2005). Can language acquisition give children a point of view? In J. Astington, & J. Baird, *Why language matters for theory of mind?* (pp. 186–219). Oxford: Oxford Press.

de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence, 199-200* , 67-92.

de Weerd, H., Verbrugge, R., & Verheij, B. (2014). Theory of Mind in the Mod Game: An Agent-Based Model of Strategic Reasoning. *Proceedings of the European Conference on Social Intelligence, 1283* , pp. 128-136.

Dennett, D. C. (1978). Beliefs about beliefs. *Brain and Behavioral Sciences, 1*, 568-570.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey, & R. Gelman (Eds.). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Doherty, M. J. (2009). *Theory of Mind: How Children Understand Others' Thoughts and Feelings.* Hove, East Sussex: Psychology Press.

Duch, W., Oentaryo, R. J., & Pasquier, M. (2008). Cognitive Architectures: Where do we go from here? In B. G. Pei Wang (Ed.), *In Proceedings of the 2008 conference on Artificial General Intelligence. 122-136*, pp. 122-136. IOS Press. Amsterdam, Netherland: IOS Press.

Epstein, J. (2012). *Generative Social Science: Studies in Agent-Based Computational Modelling.* Princeton University Press.

Epstein, J., & Axtell, R. (1996). *Growing artificial societies, social sciences from bottom up.* Washington, D.C: The MIT Press.

Farrar, M., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language, 22*, 197–213.

Flavell, J., Everett, B., Croft, K., & Flavell, E. .. (1981). Young children's knowledge about visual perception: further evidence for the level 1 level 2 distinction. *Dev Psychol, 17*(1), 99–103.

Fliss, R., Le Gall, D., Etcharry-Bouyx, F., Chauviré, V., Desgranges, B., & Allain, P. (2015). Theory of Mind and social reserve: Alternative hypothesis of progressive Theory of Mind decay during different stages of Alzheimer's disease. *Social neuroscience*, 1-15.

Frank, C., Baron-Cohen, S., & Ganzel, B. L. (2015). Sex differences in the neural basis of false-belief and pragmatic language comprehension. *NeuroImage, 105*, 300-311.

Freeman, N. H., Lewis, C., & Doherty, M. J. (1991). Preschoolers' grasp of a desire for knowledge in false-belief prediction: practical intelligence and verbal report. *British Journal of Developmental Psychology, 9*, 7-31.

Freeman, N., & Lacohee, H. (1995). Making explicit 3-year-old's implicit competence with their own false beliefs. *Cognition, 56*, 31- 60.

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B, 367*, 2213–2223.

Frith, U. (2001). Mind blindness and the brain in Autism. *Neuron*(32), 969-979.

Frith, U., Happé, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development, 3*(2), 108–124.

Gallagher, H. L., & Frith , C. (2003). Functional imaging of 'theory of mind'. *Trends in cognitive sciences, 7*(2), 77–83.

German, T. P., & Leslie, A. M. (2000). *Attending to and learning about mental states In P. Mitchell, & K. Riggs (Eds.), Children's reasoning and the mind.* Hove: Psychology Press.

Gilbert, N. (1996). Holism, individualism and emergent properties. An Approach from the Perspective of Simulation in Hegselmann, R., Mueller, U. and Troitzsch, K. G. (Eds),. *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View. Kluwer, Dordrecht*, 1-27.

Gilbert, N. (2006). Emerging Artificial Societies Through Learning,. *Journal of Artificial Societies and Social Simulation, 9*(2).

Gilbert, N. (2007). *Agent-Based Models.* London: Sage Publications.

Gilbert, N., & Troitzsch, K. G. (2011). *Simulation for the Social Scientist.* Maidenhead, Berkshire: Open University Press.

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68*(1), 70-83.

Happe, F., Winner, E., & Brownell, H. (1998). The getting of wisdom: theory of mind in old age. . *Developmental Psychology, 34*(2), 358 – 362.

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition, 101*(3), 495–514.

Hatnaa, E., & Benensonb, I. (2012). The Schelling Model of Ethnic Residential Dynamics: Beyond the Integrated - Segregated Dichotomy of Patterns. *Journal of Artificial Societies and Social Simulation , 15 (1)* (6).

Hedger, J. A., & Fabricius, W. V. (2011). True Belief Belies False Belief: Recent Findings of Competence in Infants and Limitations in 5-Year-Olds, and Implications for Theory of Mind Development. *Review of Philosophy and Psychology, 2*(3), 429-447.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: models, analysis and simulation. *Journal of Artificial Societies and Social Simulation, 5*(3).

Hollebrandse, B., Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning. *Synthese, 191*(3), 321-333.

Hughes, C. (1998). Executive function in preschoolers: links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.

Hughes, C. (2011-a). *Social Understanding and Social Lives: From Toddlerhood through to the Transition to School.* Hove: Psychology Press.

Hughes, C. (2011-b). Changes and challenges in 20 years of research into the development of executive functions. *Infant and Child Development, 20*(3), 251–271.

Hughes, C., Ensor, R. A., Allen, L. L., Devine, R. T., De Rosnay, M., Koyasu, M., & . . . Lecce, S. (2011). Theory of mind performance in British, Australian, Japanese and Italian children: Contrasts in culture or age of school entry? *Paper presented at the Society for Research in Child Development (SRCD) Biennial Conference, Montreal, Canada*.

Humphrey, N. (1976). The social function of intellect. Bateson P.P.G. *In Growing points in ethology, ed. P.P.G.Bateson and R.A.Hinde*, 303–317.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition, 140*, 14–23.

Jones, A., Gutierrez, R., & Ludlow, A. (2015). Confronting the language barrier:Theory of mind in deaf children. *Journal of Communication Disorders, 56* , 47–58.

Jones, M. T. (2008). *Artificial Intelligence: A Systems Approach* . Sudbury, MA: Jones and Bartlett Publishers.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109* , 224–234.

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2016). Differing views: Can chimpanzees do Level 2 perspective-taking? *Animal Cognition, 19*(3), 555-564.

Keenan, T. (1998). Memory span as a predictor for false belief understanding. *New Zealand Journal of Psychology, 27*(2), 36–43.

Keenan, T., Olson, D. R., & Zopito, M. (1998). Working Memory and Children's Developing Understanding of Mind. *Australian Journal of Psychology, 50*(2), 76–82.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330* (6012), 1830–1834.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science, 12* (4), 521–535.

Laird, J. E., & Bates Congdon, C. (2015). *The Soar User's Manual. Version 9.5.0.* University of Michigan.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research , 10*(2), 141-160.

Leslie, A. (2005). Developmental parallels in understanding minds and bodies. *Trends in cognitive sciences,, 9*(10), 459–462.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 2*(1), 247–253.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*(12).

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*(1), 45-85.

Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: conceptual deficit or linguistic artifact? *Child Development, 61*, 1514-1519.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). Twelve- and 18-month-olds point to provide information for others. *Journal of Cognition and Development, 7*(2), 173–187.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of Mind Development in Chinese Children: A Meta-Analysis of False-Belief Understanding Across Cultures and Languages. *Developmental Psychology, 44*(2), 523–531.

MacLeod, C. M. (2007). *Inhibition In Cognition* (Vol. xvii). (D. S. Gorfein, Ed.) Washington, DC, USA: American Psychological Association.

Marcovitch, S., O'Brien, M., Calkins, S., Leerkes, E., Weaver, J., & Levine, D. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development*.

Marsella, S. C., Pynadath, D. V., & Read, S. J. (2004). PsychSim: Agent-based modeling of social interactions and influence. *Proceedings of the international conference on cognitive modeling., 36*, pp. 243-248.

Martin, A., & Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition, 130*, 300–308.

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences, 20*(5), 375–382.

Milligan, K., Astington, J., & Dack, L. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646.

Minsky, M. (1988). *The Society of Mind.* New York: Simon & Schuster.

Mitchell, J., Macrae, C., & Banaji, M. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *The Journal of Neuroscience, 24*(21), 4912-7.

Mitchell, P. (1996). *Acquiring a conception of mind : a review of psychological research and theory*. Hove : Psychology Press.

Mitchell, P., & Lacohee, H. (1991). Children's early understanding of false belief. *Cognition, 39*, 107-127.

Mohnke, S., Erk, S., Schnell, K., Romanczuk-Seiferth, N., Schmierer, P., Romund, L., . . . Walter, H. (2015). Theory of mind network activity is altered in subjects with familial liability for schizophrenia. *Social Cognitive and Affective Neuroscience*.

Moll, H., & Meltzoff, A. N. (2011). How Does It Look? Level 2 Perspective-Taking at 36 Months of Age. *Child Development, 82*(2), 661–673.

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B,, 362*(1480), 639-648.

Moses, L. (2001). Executive accounts of theory of mind development. *Child Development, 3*, 688-690.

Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development, 8*, 1-25.

Onggo, B. S., & Karpat, O. (2011). Agent-based conceptual model representation using BPMN. *In Proceedings of the 2011 Winter Simulation Conference, S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu (ed.)*, (pp. 671-682).

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255-258.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Perner, J., & Leekam, S. .. (2008). The curious incident of the photo that was accused of being false: issues of domain specificity in development, autism, and brain imaging. *The Quarterly Journal of Experimental Psychology, 61* (1), 76–89.

Peterson, C., Garnett, M., Kelly, A., & Attwood, T. (2009). Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *European Child and Adolescent Psychiatry, 18*, 105–115.

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98*(2), 137-155.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based Simulation Platforms: Review and Development Recommendations. *Simulation, 82*(9), 609-623.

Roekel, E. v., Scholte, R. H., & Didden, R. (2010). Bullying Among Adolescents With Autism Spectrum Disorders: Prevalence and Perception. *Journal of Autism and Developmental Disorders, 40*(1), 63–73.

Rubio-Fernández, P. (2013). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science, 24*(1), 27-33.

Russell, J. (1996). *Agency: Its role in mental development.* Hove: Erlbaum (UK), Taylor & Francis.

Russell, S., & Norvig, P. (2014). *Artificial Intelligence: A Modern Approach.* Essex: Pearson.

Salamon, T. (2011). *Design of Agent-Based Models.* Czech Republic: Tomas Bruckner .

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective self-perspective. *Brain,* 1102-1111.

Santos, L., Nissen, A., & Ferrugia, J. (2006). Rhesus monkeys (Macaca mulatta) know what others can and cannot hear,. *Animal Behaviour, 71* , 1175–1181.

Schaafsma, S., Pfaff, D., Spunt, R., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences, 19*(2), 65-72.

Schelling, T. C. (1971). Dynamic Models of Segregation. *The Journal of Mathematical Sociology, 1*, 143-186.

Schelling, T. C. (2006). *Micromotives and Macrobehavior.* New York, USA: W.W. Norton & Company, Inc.

Schmelz, M., Call, J., & Tomasello, M. (2011). Chimpanzees know that others make inferences. *Proceedings of the National Academy of Sciences , 108*(7), 3077–3079.

Scholl, B. J., & Leslie, A. (2001). Minds, modules, and meta-analysis. *Child Development, 72*(3), 696-701.

Schunn, C., & Gray, W. (2002). Introduction to the special issue on computational cognitive modeling. *Cognitive Systems Research, 3*(1), 1-3.

Scotta, R. M., Baillargeon, R., Song, H.-j., & Leslie, A. (2010). Attributing false beliefs about non-obvious properties at. 18 months. *Cognitive Psychology, 61*(4), 366-395.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38*(1), 1-12.

Simon, H. A. (1957). *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* New York: John Wiley and Sons.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory-of-mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23(1)*, 117 – 141.

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science, 10*(2), 199-204.

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition, 130*, 1-10.

Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation Through Attribution of False Belief by 2-Year-Olds. *Psychological Science, 18*, 587–592.

Sun, R. (2006). *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation.* New York: Cambridge University Press.

Sun, R. (2008). *The Cambridge Handbook of Computational Psychology.* New York: Cambridge University Press.

Sun, R. (2012). *Grounding Social Sciences in Cognitive Sciences* . Cambridge, Massachusetts : The MIT Press.

Sun, R., Coward, L. A., & Zenzen, M. J. (2005). On levels of cognitive modeling. *Philosophical Psychology, 18* (5), 613-637.

Surian, L., & Leslie, A. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology, 17*, 141–155.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580–586.

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science, 16*(6), 311-315.

Tetlock, P., & Goldgeier, J. (2000). Human nature and world politics: cognition, influence, and identity. *International Journal of Psychology, 35*, 87–96.

Unified Modeling Language (UML). (2015). Retrieved from OMG: http://www.omg.org/spec/UML/

Vaart, E. v., & Hemelrijk, C. K. (2014). 'Theory of mind' in animals: ways to make progress. *Synthese, 191*, 335-354.

Verbrugge, R. (2009). Logic and Social Cognition. The Facts Matter, and so do Computational Models. *Journal of Philosphical Logic, 38*, 649–680.

Volkmar, F., Sparrow, S. G., Cicchetti, D., Paul, R., & Cohen, D. (1987). Social deficits in autism: an operational approach using the Vineland Adaptive Behaviour Scales. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*(2), 156–161.

Wang, L., & Leslie, A. M. (2016). Is Implicit Theory of Mind the 'Real Deal'? The Own-Belief/True-Belief Default in Adults and Young Preschoolers. *Mind & Language, 31*(2), 147-176.

Watson, A. C., Painter, K. M., & Bornstein, M. H. (2001). *Journal of Cognition and Development, 2*(4), 449-457.

Weiss, G. (2013). *Multiagent systems (2nd ed.).* Cambridge, MA: The MIT Press.

Wellman, H. (1990). *The Child's Theory Of Mind.* (MA): MIT Press.

Wellman, H. (2014). *Making Minds. How Theory of Mind Develops.* NewYork: Oxford university Press.

Wellman, H. M. (1992). *Do children have a theory of mind?* Cambridge: The MIT Press.

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*, 239-277.

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

Whiten, A. (2000). Social complexity and social intelligence . In N. Foundation, *The Nature of Intelligence.* John Wiley & Sons,.

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NETLogo.* London: MIT Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition , 13*(1), 103–128.

Wimmer, W., & Hartl, H. (1991). Against the Cartesian view on mind: Young children's difculty with own false beliefs. *British Journal of Developmental Psychology, 9*, 125-128.

Wooldridge, M. (2009). *Introduction to MultiAgent Systems.* West Sussex: John Wiley & Sons.

Workman, L., & Reader, W. (2014). *Evolutionary Psychology: An Introduction.* New York: Cambridge University Press.

Yazdi, A., German, T., Defeyter, M., & Siegal, M. (2006). Competence and performance in belief-desire reasoning across two cultures: the truth, the whole truth and nothing but the truth about false belief? *Cognition, 100*(2), 343-68.

Yilmaz, L. (2015). *Concepts and Methodologies for Modeling and Simulation: A Tribute to Tuncer Ören.* New York: Springer.

APPENDIX 1

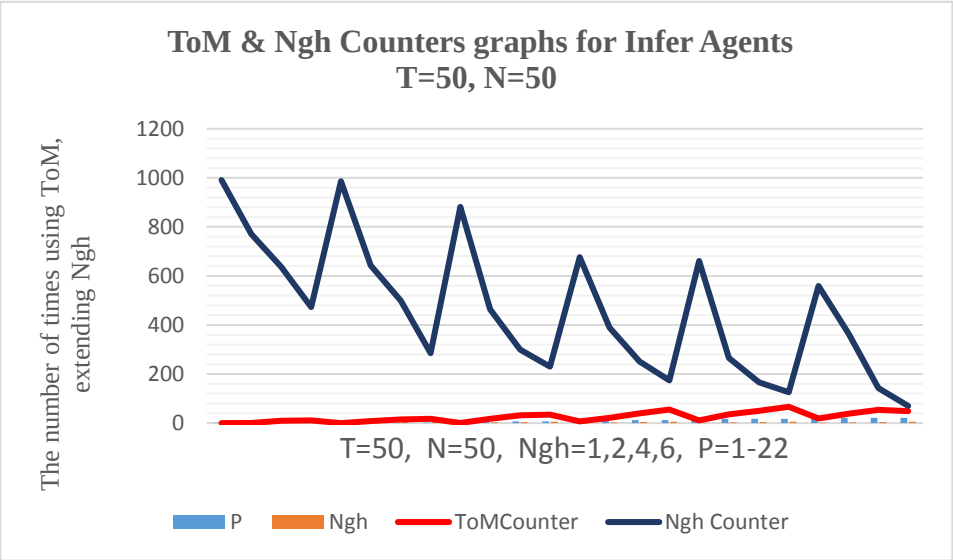


Figure 95. The number of times ToM and Ngh functions are used: T=50, N=50

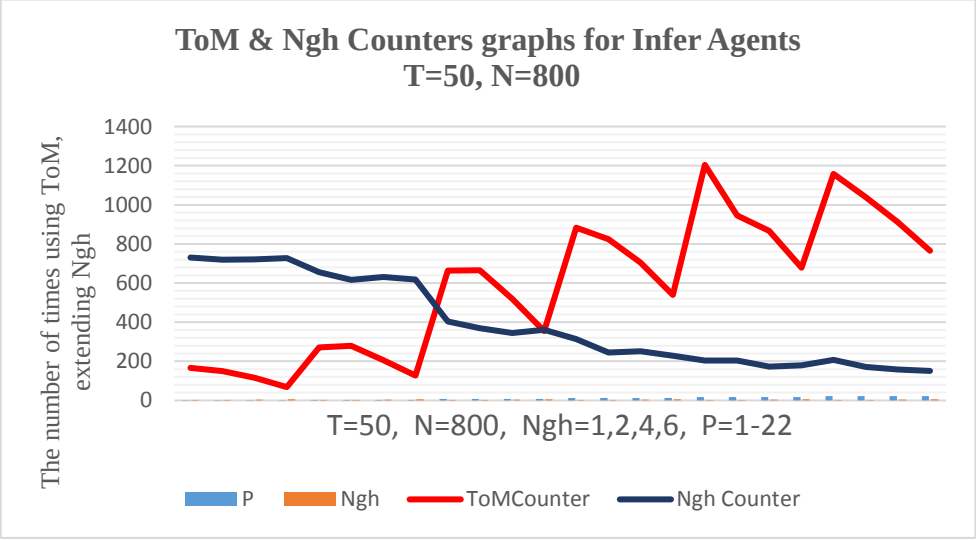


Figure 96. The number of times ToM and Ngh functions are used: T=50, N=800

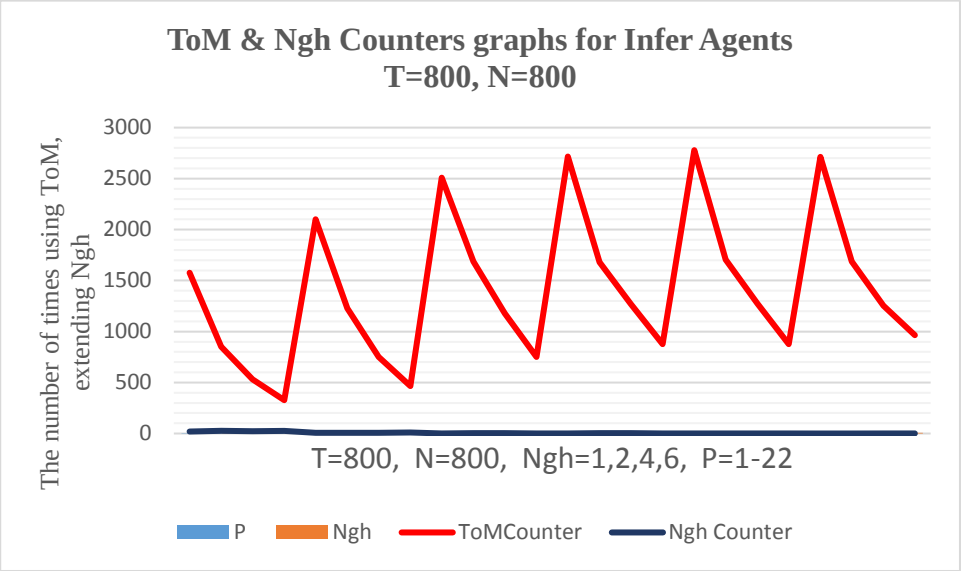


Figure 97. The number of times ToM and Ngh functions are used: T=800, N=800

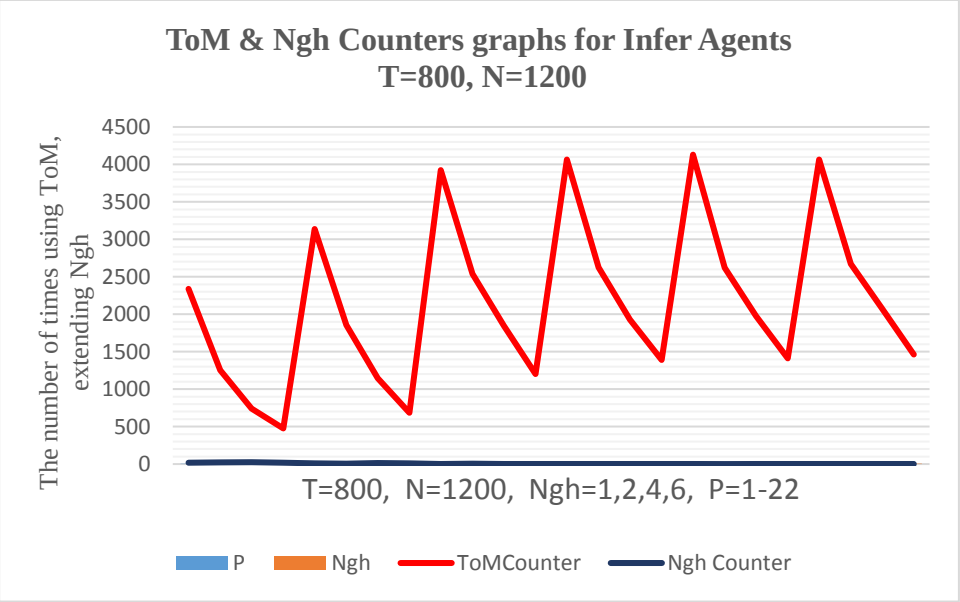


Figure 98. The number of times ToM and Ngh functions are used: T=800, N=1200

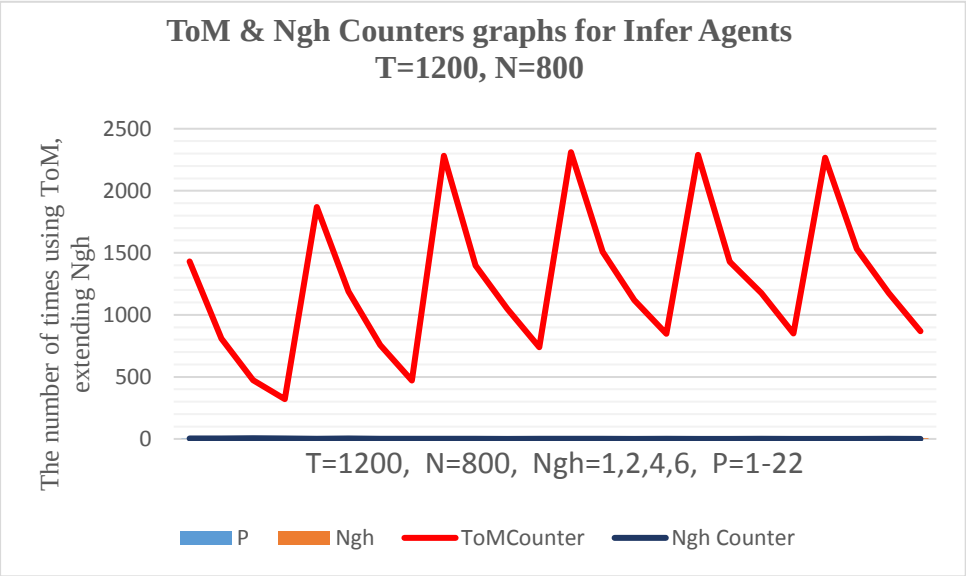


Figure 99. The number of times ToM and Ngh functions are used T=1200, N=800

APPENDIX 2

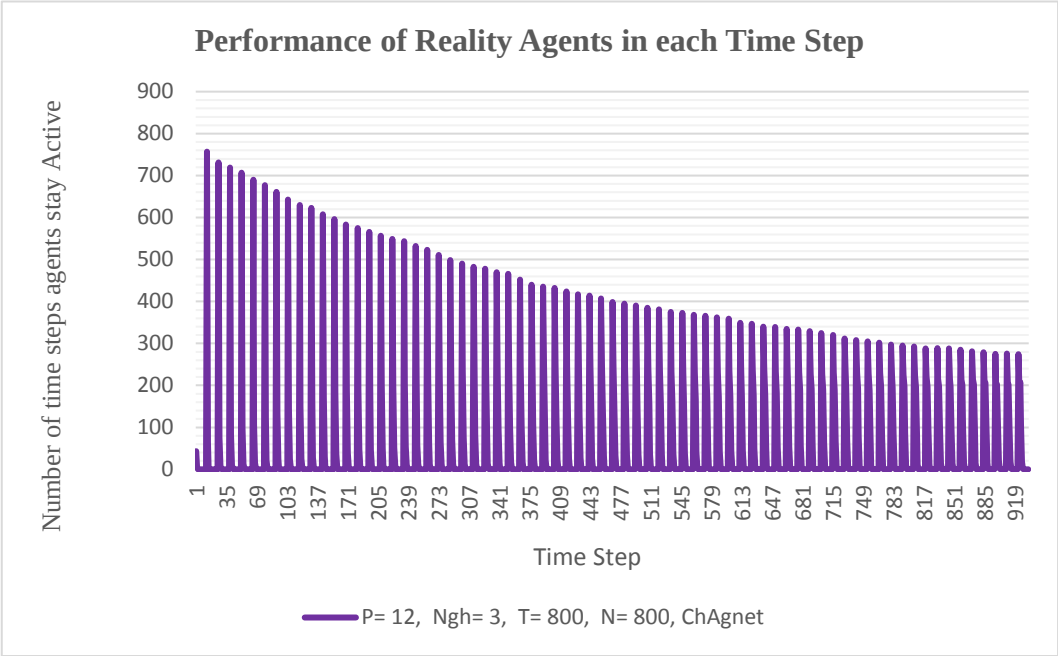


Figure 100. Performance of Reality Agents in each Time Step

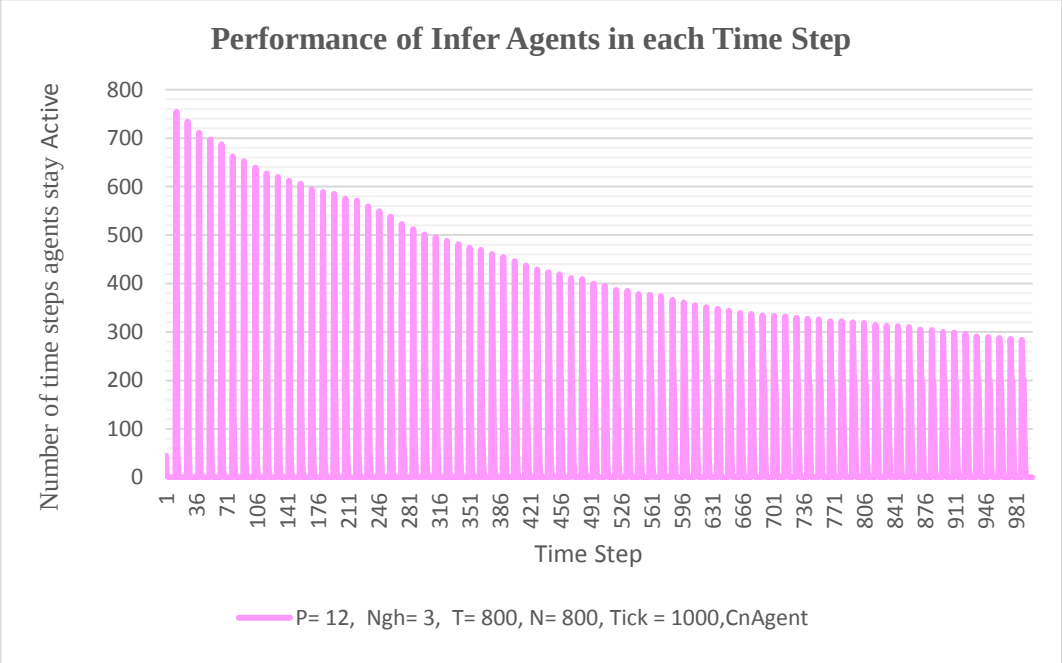


Figure 101. Performance of Infer Agents in each Time Step

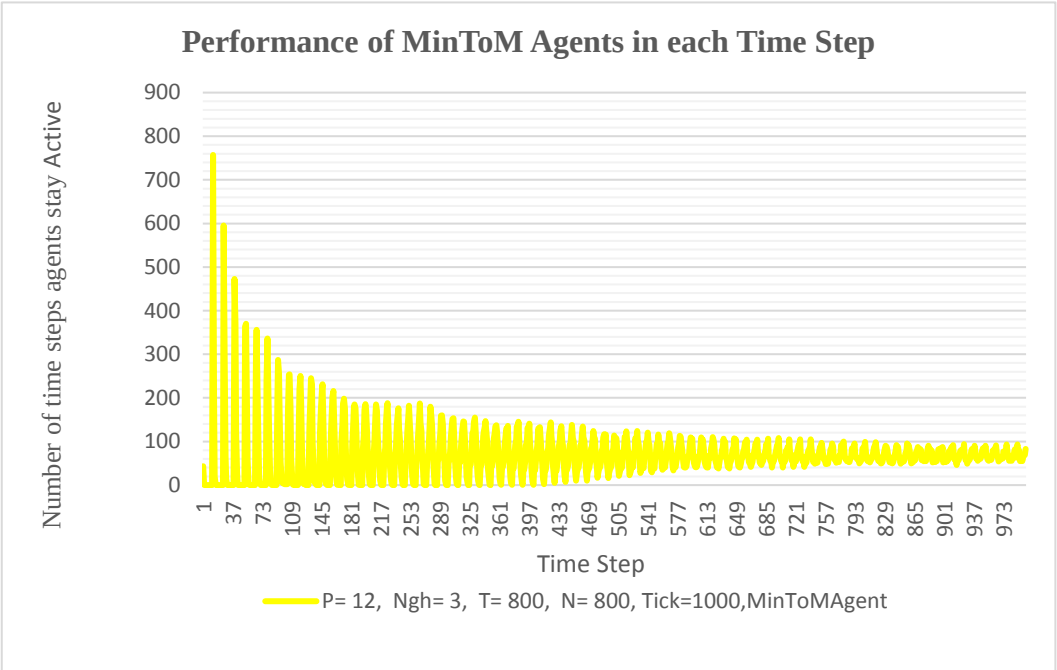


Figure 102. Performance of MinToM Agents in each Time Step

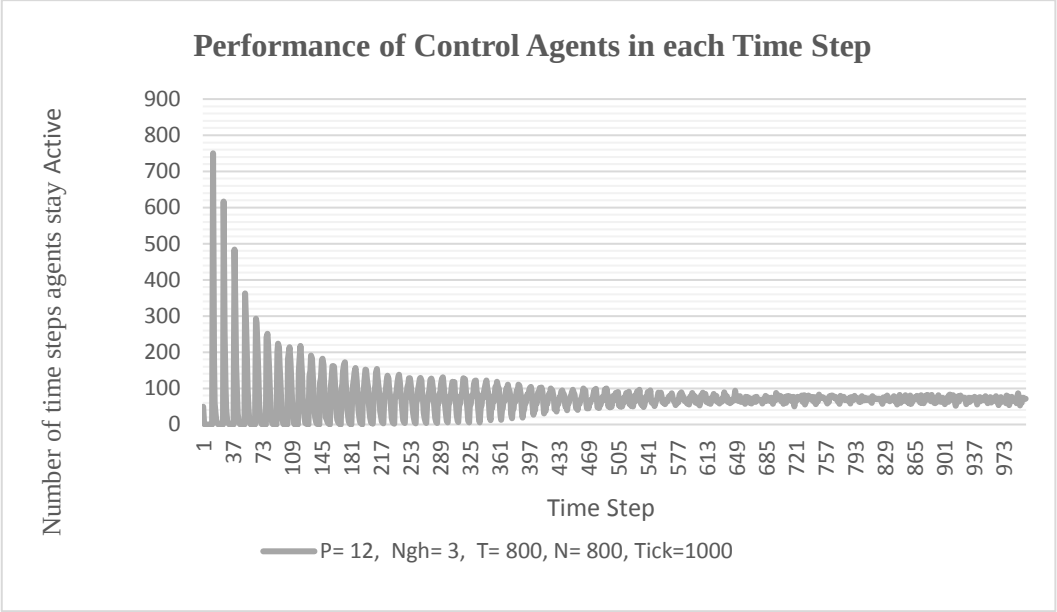


Figure 103. Performance of Control Agents in each Time Step

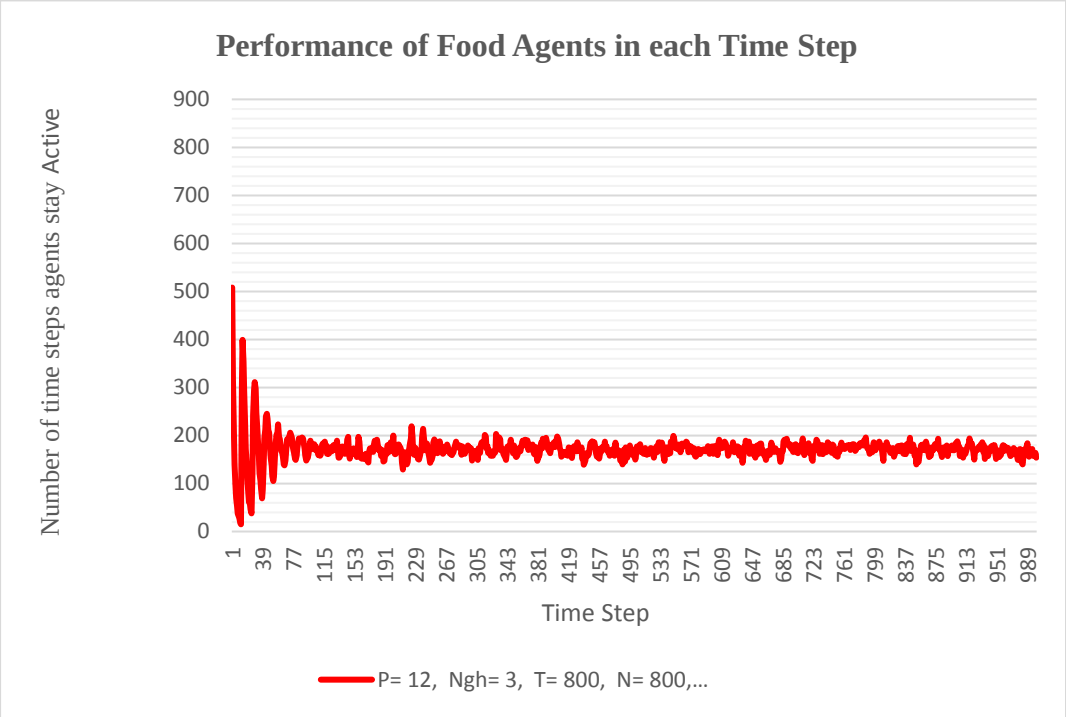


Figure 104. Performance of Food Agents in each Time Step

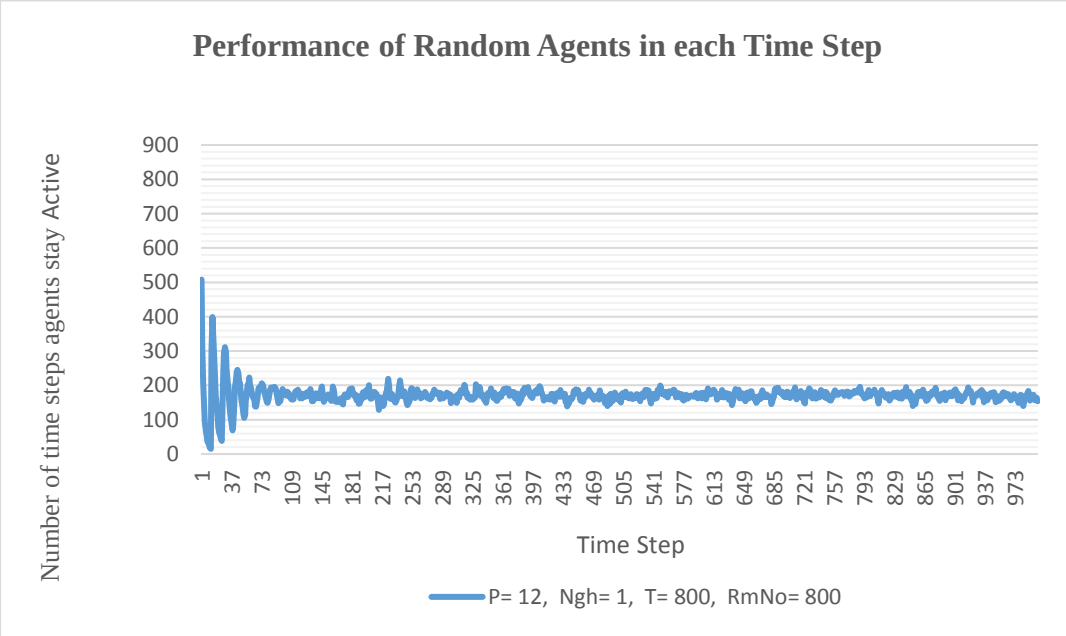


Figure 105. Performance of Random Agents in each Time Step