university of groningen

faculty of science and engineering

Smart Negotiators for a Smarter Grid: Exploring the Potential of Theory of Mind in Energy Regulation

Isabelle Tilleman

Master Thesis
Artificial Intelligence
University of Groningen, The Netherlands

November 2023

Internal Supervisor:
Prof. dr. Rineke Verbrugge (Artificial Intelligence, University of Groningen)

External Supervisor:
Dr. Wico Mulder (TNO, Groningen)

university of groningen / faculty of science and engineering

University of Groningen

To fulfil the requirements for the degree of
Master of Science in Artificial Intelligence
at the University of Groningen under the supervision of
prof. dr. Rineke Verbrugge (Artificial Intelligence, University of Groningen)
and
dr. Wico Mulder (TNO, Groningen)

Isabelle Tilleman (s3656586)

November 14, 2023

Contents

Acronyms

BB Bernoulliborg

LB Linnaeusborg

ToM Theory of Mind

HVAC Heating, Ventilation and Air Conditioning

ToM_0 zero-order Theory of Mind

ToM_1 first-order Theory of Mind

ToM_2 second-order Theory of Mind

ToM_3 third-order Theory of Mind

ToM_k k-th other Theory of Mind

Acknowledgements

This thesis would not be what it is today without the help of several others.

First and foremost, I would like to thank my supervisors Rineke Verbrugge and Wico Mulder. Both of you have been very helpful during the process of writing this thesis. Rineke, thank you for your detailed feedback and for helping me define the scope of my work. Wico, thank you for your enthusiasm and for always keeping me on my toes. I feel extremely lucky to have had the both of you as my supervisors.

Second, I would like to thank everyone at TNO: the department of Data Ecosystems, the wonderful people at the Groningen office, and of course my fellow interns! You offered me advice when I needed it, and helped me get my mind off of everything when I needed to be able to take a step back.

Third, I would like to thank everyone who kept me motivated throughout the process. While this also holds for the people mentioned above, I would also like to thank my partner and my friends, who motivated me to work when I was not at TNO and gave me words of support when I needed it.

Fourth, I would like to thank everyone who helped me shape the ideas in this thesis. Specifically, I would like to thank J.D. Top, Aliene van der Veen, and the people at AIMZ for the insightful discussions we had during my time working on this thesis.

Lastly, I would like to thank Harmen de Weerd, who did all the work that came before. Thank you Harmen, for all the work you have done on the topic of theory of mind, for your guidance, and for providing me with your code.

Abstract

Theory of mind is the ability to attribute unobservable mental content to others [25]. This ability has been shown to provide an advantage in cooperative, competitive, and mixed-motive settings [34]. Previous research on the advantages of using theory of mind has mainly been limited to theoretical and fundamental research.

The goal of this thesis was to extend previous findings about theory of mind into the applied domain, specifically that of regulating energy consumption between buildings in close proximity to prevent overloading the energy grid, by answering the following research question: "How can theory of mind be used in computational models of negotiations in a Human-AI ecosystem, to be tested by way of a game?".

To answer this question, we also answered three sub-questions: "How can the Coloured Trails game be modified to correspond with real-life negotiations about energy consumption between two buildings?", "How can the modified Coloured Trails be used to create a multi-agent system that simulates negotiations about energy consumption between two buildings?", and "Does using theory of mind provide an advantage in simulated negotiations about energy consumption between two buildings?".

To answer the first sub-question, we modified the Coloured Trails game to correspond with real-life negotiations about energy consumption to create a new game called Energy Trails. This novel adaptation of Coloured Trails features multiple playing boards, the possibility to prioritise between the goals on different boards, and a different kind of goal for players: their goal is now to reach a certain composition of their trail on each board, rather than reaching a certain location on the board. Additionally, the scoring was adjusted to create two versions of Energy Trails: one with competitive scoring, which is similar to the scoring in Coloured Trails, and one with cooperative scoring, in which players can get also points if their trading partner reaches a goal.

To answer the second sub-question, we created a multi-agent simulation of this new adaptation of Coloured Trails. For this, the simulation of Coloured Trails and theory of mind agents able to play it by DeWeerd [34] was used as a foundation. The simulation of Coloured Trails was turned into a simulation of Energy Trails. The agents stayed largely the same, but needed some adjustments in their reasoning to deal with the increased complexity of Energy Trails compared to Coloured Trails. These adjustments consisted of shortcuts in the belief-updates for internal simulations of the player's trading partners.

To answer the third sub-question, experiments were performed across four conditions: 2 boards with competitive scoring, 2 boards with cooperative scoring, 4 boards with competitive scoring, and 4 boards with cooperative scoring. In the conditions with 2 boards, the original reasoning of theory of mind agents, without shortcuts, was used.

The performance of theory of mind agents was compared to that of agent pairings with exclusively zero-order theory of mind agents based on the following metrics: acceptance rate, mean score increase, mean score increase after offer acceptance, mean number of negotiation rounds used before the negotiation was ended, and time-out rate. The overall performance of theory of mind agents was validated by considering the percentage of negotiations in which the goal profile they believed to most likely be that of their trading partner was indeed the goal profile of their trading partner. This percentage was compared to the expected percentage for agents which guess randomly to check whether our agents could outperform a random guesser.

The results of these experiments showed the following effects: In terms of acceptance rate, mean score increase across all negotiations, and number of negotiation rounds, the conditions with the competitive scoring most consistently showed an advantage for theory of mind agents. In the cooperative scoring, this advantage was only consistently found for agent pairings in which no zero-order theory of mind agents were present. When it came to the mean score increase after an offer was accepted, the advantage was most clear in the conditions that used cooperative scoring. The performance of theory of mind agents in terms of belief correctness was consistently better than that of a random guesser across all conditions for all agent pairings.

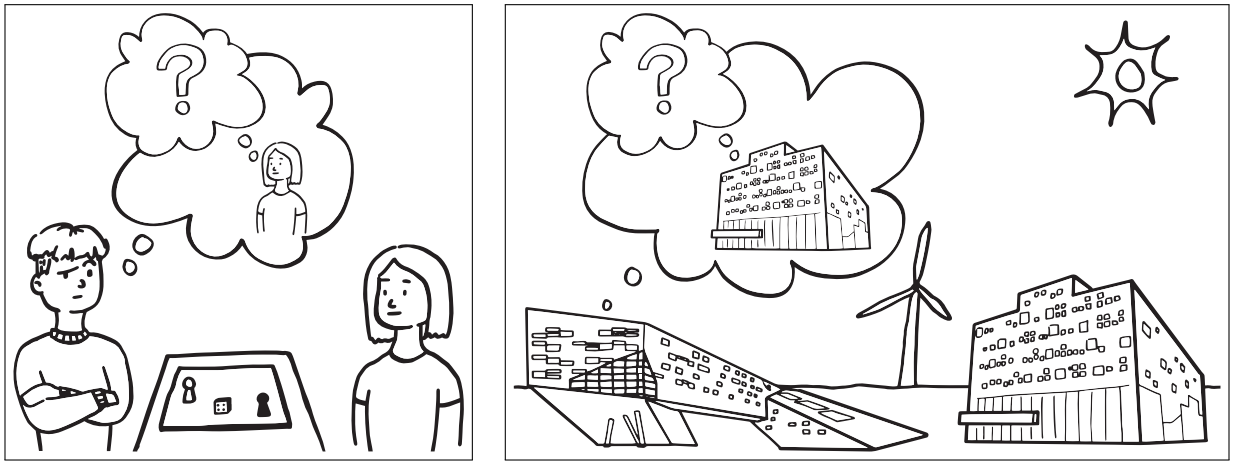
Improvements that can be made to the work done in this thesis include improved parameter tuning, preventing negotiation cycles, and finding better ways to deal with how the increased complexity of Energy Trails affects theory of mind reasoning. This thesis was able to bridge create a bridge between fundamental theory of mind research and the application of theory of mind in the energy domain. Potential future work includes generalising the Energy Trails game to make it applicable to more domains, using the Energy Trails game to teach humans how to negotiate about energy consumption using theory of mind, and implementing Energy Trails and theory of mind-based negotiations into systems that handle energy regulation in real buildings.

To conclude, we found that theory of mind can be used in computational models of negotiations in a Human-AI ecosystem by letting agents negotiate in the Energy Trails game. This was confirmed by testing the performance of theory of mind agents in a multi-agent simulation of the Energy Trails game. The results of these experiments show that theory of mind agent pairings always perform equal to or better than zero-order theory of mind agent pairings in the competitive version of Energy Trails. If no zero-order agents are present in the negotiation, this finding can be extended to both the competitive and cooperative versions of Energy Trails.

1 Introduction

Do you like strategic games? Then you might be familiar with the following scenario: You try to figure out what your opponent might do next, but then you get stuck. What if they do not do what you expect them to do? Maybe, this happens exactly because they expect you to expect them to do the expected. So instead, they will do the unexpected. But then, what if they also expect you to expect the unexpected, so they will opt to do the expected instead?

Sounds complicated? It is! This type of reasoning is an example of Theory of Mind (ToM), the ability to reason about the unobservable mental content of others [25], which is often used to figure out what someone else's goals and beliefs might be. The example above explains how theory of mind is used in a competitive setting. A small illustrated example of theory of mind in humans and of how we want to use theory of mind in this study can be seen in Figure 1. A more extensive explanation of theory of mind is given in Chapter 2.



(a) An illustrative example of theory of mind in humans.

(b) An illustrative example of theory of mind in buildings.

Figure 1: Two illustrative examples of theory of mind in humans and in buildings.

Previous research has shown in which settings theory of mind can provide an advantage, and found that it is useful in competitive, cooperative, and mixed-motive settings [36]. This shows that theory of mind is versatile and applicable to many situations. Yet, there is not much research about how theory of mind can be used to solve real-life problems, in particular, in the domain of division of scarce resources.

We have seen an example of how we can use theory of mind in a competitive setting. As explained, we can also use it in a cooperative setting. Consider the following situation: You have an appointment with your supervisor, but you never specified where the meeting would take place, and your phone is out of battery. Since you cannot specify where the meeting will take place, or check if your supervisor has expressed a preference, you need to attempt to figure out which option is the most likely. Will they want to meet at their office, online, or maybe at a third location? In turn, your supervisor is likely to go through the same thought process, since they have not heard from you. In this situation, the inability to communicate forces both of you to use theory of mind to reach the goal of meeting your supervisor.

In order to understand how theory of mind works in a mixed-motive setting, which is neither purely cooperative nor purely competitive, we need to consider a different example: You and your partner are trying to distribute the house chores. This is a negotiation of sorts, in which you do want to end up with a fair distribution of tasks, but you would also like to end up with the chores you like best, you need to work together, compromise, and negotiate. In this negotiation, it can help to consider what you expect your partner to want. If you know they like doing the dishes, and you like doing the laundry, you can offer to do the laundry if they will do the dishes. This way, your partner is likely to let you do the laundry because they will also get a chore they enjoy doing. Here, theory of mind helps you reach a better division of tasks as a result of negotiating, faster.

In today's society, sustainability and regulation of resources are an important topic. Fossil energy is expected to run out in the next 100 years [24] and renewable energy resources may not be enough unless we learn to deal with lower consumption rates [29, 33].

This requires work on our part. A good first step to take towards a future in which we can expect the per-capita energy consumption levels to be lower, is to find ways to distribute the limited energy that is available. Since this is a complex problem that humans may not always be equipped to solve, we can try to create systems that deal with this problem instead.

We will look at the following example: Let us consider two buildings on the Zernike Campus in Groningen. From the buildings on the Zernike Campus, we chose two to be the main protagonists of our example: the Bernoulliborg (BB) and the Linnaeusborg (LB). Both of these buildings are managed by the University of Groningen, specifically by the Faculty of Science and Engineering. These buildings can be seen in Figure 2.

(a) Picture of the Bernoulliborg: Nijenborgh 9, 9747 AG Groningen.

(b) Picture of the Linnaeusborg: Nijenborgh 7, 9747 AG Groningen. ©UG, photo: Silvio Zangerini

Figure 2: Pictures of the buildings used for the example scenario.

Since these two buildings are situated close to each other, we assume that they use the same energy grid. They need to coordinate how much energy they consume during different times of day in order to not overload the grid. Since coordinating their energy consumption levels would require them to collaborate whilst trying to reach their own personal goals, such as providing a comfortable working temperature for employees in the building, we consider it to be a mixed-motive setting, just like the house chore example we saw earlier.

From earlier work, we know that using theory of mind can be beneficial in mixed-motive settings [34]. In order to find out whether this finding also extends to real-life settings, we decided to investigate how effective theory of mind is in a simulation of the building context described above and as illustrated in Figure 1a.

This leads us to the following research question:

> "How can theory of mind be used in computational models of negotiations in a Human-AI ecosystem, to be tested by way of a game?"

In order to answer this question, we have divided it into sub-questions.

Since we know that using theory of mind was found to improve performance in the Coloured Trails game [34], which is a mixed-motive negotiation game, we decided to use this game as a basis for our model of the earlier described problem in the energy domain. This has led to the following sub-question:

> "How can the Coloured Trails game be modified to correspond with real-life negotiations about energy consumption between two buildings?"

In order to stay close to earlier work [36], we decided to use a similar set-up to create a computational model of negotiations. This led to the following sub-question:

> "How can the modified Coloured Trails be used to create a multi-agent system that simulates negotiations about energy consumption between two buildings?"

Lastly, to verify whether this approach is worth the effort for parties who would consider using a theory of mind based solution in their energy regulation systems, we also answered the following sub-question:

> "Does using theory of mind provide an advantage in simulated negotiations about energy consumption between two buildings?"

The rest of this Master's Thesis is divided in the following sections. Chapter 2 provides an overview of the previous research and literature relevant to the work done in this thesis and how they led to the aforementioned research questions. Chapter 3 contains the answers to the first two sub-questions by describing our newly created variation of Coloured Trails called Energy Trails in Section 3.1, as well as describing the simulation that was built in Section 3.2. Chapter 4 contains the results of the experiments as outlined in Section 3.3. Chapter 5 contains the overall findings of this research, as well as an explanation of the scientific relevance, improvements that can be made, and future research.

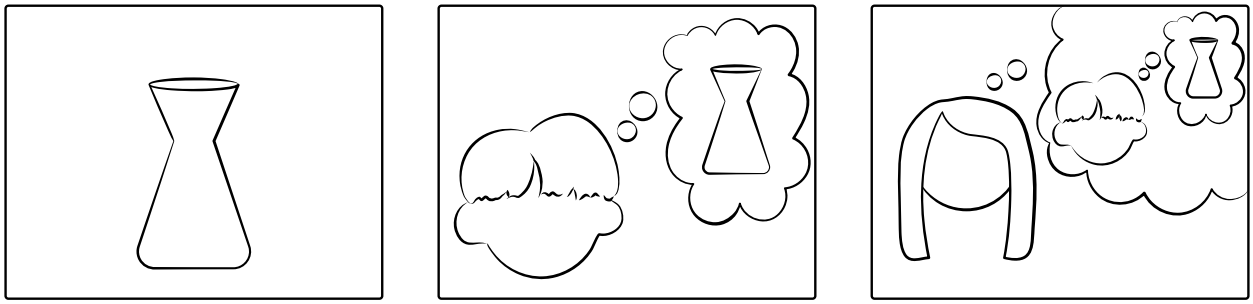
2 Background Literature

This chapter will provide a comprehensive overview of the theoretical knowledge that was relevant during this research, particularly in the domains of theory of mind, energy regulation, and negotiation and game theory.

2.1 Theory of Mind

Theory of Mind (ToM), a term coined in 1978, is the ability to attribute mental states which are otherwise unobservable to others [25]. This ability can be used to infer, for example, someone else's intentions, goals, or beliefs. When someone does not possess the ability to use theory of mind, they can only observe an outcome. Such a person or agent can also be said to possess Zero-order Theory of Mind (ToM_0) [34]. This means they can only reason about their own mental state.

Theory of mind is recursive. The order of theory of mind denotes the recursive limit that is put on it. To illustrate the different orders of theory of mind, we will consider a fictional scenario as seen in Figure 3. ToM_0 is equivalent to the lack of theory of mind, which means such an agent can only reason about what it can observe directly, such as the statement "The vase is empty" which is illustrated in Figure 3a. First-order Theory of Mind (ToM_1) allows an agent to reason about the mental states of others, such as the attribution "Sam knows that the vase is empty" which is illustrated in Figure 3b. Second-order Theory of Mind (ToM_2) and above allows an agent to reason about another agent's theory of mind, such as the attribution "Rachel knows that Sam knows that the vase is empty" which is illustrated in Figure 3c. The ability to reason about theory of mind itself is available to anyone with ToM_2 or higher, and is referred to as higher-order theory of mind.



(a) A ToM_0 attribution example. (b) A ToM_1 attribution example. (c) A ToM_2 attribution example.

Figure 3: Three example scenarios to illustrate which types of attributions the different orders of ToM agents can reason about

There are multiple perspectives on theory of mind [30, 34]: theory-theory of mind [17, 19] and simulation-theory of mind [2]. Theory-theory of mind argues for a rule-based system, in which there is an understanding of how the mind works. Simulation-theory of mind argues for a system where one's own mental state is used as an approximator for the state-of-mind of another. According to this theory, since the mental state of another cannot simply be described through a set of rules, it can be seen as something like a black-box. This creates the need to approximate the other person's mental state by imagining what you would do in their shoes.

Since the computational model created for this research will partly be based on previous work [34], the assumption of said previous work will be used here, which is that agents make use of simulation-theory of mind. This is especially useful because we will be describing a situation in which humans and agents interact with each other. With the assumption that the explainability of non-human agents in the model is a couple of steps behind the state-of-the-art in explainable AI, these agents can be expected to be seen as a black boxes from the perspective of the humans that interact with them.

As mentioned, the model built in this study involves human interaction as well. This means that it is important to consider to which extent the idea we have of theory of mind actually applies to adult humans. It has been found that prior knowledge about a situation or outcome can negatively affect an adult's ability to reason about the mental state of others, this is known as the *curse of knowledge* [3]. More recently, this has been found to be attributable to a misattribution of fluency and common knowledge [4]. On the contrary, other studies have found that rather than knowledge, an adult's ability to reason about the mental state of others is often affected negatively by uncertainty [27]. While there is no consensus on the cause, previous research consistently shows that adults are not able to use theory of mind perfectly. Thus, we expect there to be a difference in performance between both human and non-human agents in the model, as well as between simulated and non-simulated human data.

The added benefit of using (higher-order) theory of mind depends on the circumstances in which it is used, especially when considering the large cost associated with using it. Previous work has attempted to find out why it was evolutionarily advantageous for theory of mind to develop in humans [34]. The effectiveness of theory of mind was tested in three settings, competitive, cooperative, and mixed-motive, to find out if theory of mind may have developed in order for humans to thrive in one of these three environments. The findings regarding use of theory of mind in these three settings are outlined in the next three sections.

2.1.1 Competitive Settings

The Machiavellian intelligence hypothesis states that intelligence developed because of a need to be good at deception and social manipulation [8, 41], this is referred to as Machiavellian intelligence. [1] In competitive settings like the ones described by the Machiavellian intelligence hypothesis, it is expected that theory of mind is used to understand and execute deception [34].

In previous research, there are multiple theories for the circumstances in which humans use higher-order theory of mind in competitive games. One theory is that humans use theory of mind more effectively when playing simple competitive games [16]. Where some previous research had been pessimistic about use of theory of mind overall, a study found that humans often use the highest order of theory of mind that is available to them when playing simple games [16]. Another theory is that people can be encouraged to use theory of mind through stepwise training, in which they are encouraged to make and explain decisions with increasingly higher orders of theory of mind [31].

The performance of different orders of ToM agents was tested in competitive settings [38]. One of the settings in which it was tested was zero-sum games, specifically three versions of rock-paper-scissors to test performance in single-shot games, and an adaptation of limited bidding, a game which is played over multiple rounds. The three variations of rock-paper-scissors in which the performance of theory of mind was tested are: rock-paper-scissors [22], elemental rock-paper-scissors [38], and rock-paper-scissors-lizard-spock [20]. In zero-sum games, the expected outcome is always zero, it

[1] We will use the same competitive social intelligence definition as is used in the work by de Weerd [34].

is not possible for both players to win. It was found that the performance of agents is higher when they use first- or second-order theory of mind than when they do not use theory of mind. Orders of theory of mind beyond the second order were not found to have a significant benefit. Additionally, the benefit seems greater in more complex games. This goes against the theory that simple games prompt people to use higher-order theory of mind [16]. These results were based on an assumption of rationality and predictability.

The similarity between the behaviour of ToM agents and humans was tested with the Mod game [13], which is an n-player generalisation of rock-paper-scissors. It was shown that the behaviour of higher-order ToM agents was more similar to the behaviour of humans playing the game than to the behaviour of ToM_0 and ToM_1 agents [40]. The results suggest that repeating competitive games makes people use higher orders of theory of mind. This seems to point in the direction of the theory that stepwise training encourages the use of higher-order theory of mind [31].

2.1.2 Cooperative Settings

The Vygotskian intelligence hypothesis [18, 21, 32] emphasises the cooperative aspects of intelligence, arguing that collaboration and cooperation play an important part in the development of certain human cognitive skills. Theory of mind is one of those skills. It allows the construction of joint intentions and shared goals. In turn, this improves collaboration and cooperation.

The performance of different orders of ToM agents was tested in a cooperative setting using the tacit communication game [5, 23, 26], a purely cooperative game. Humans are known to be very good at the variant of this game with which the performance of ToM agents was tested [26]. It was tested whether the agents performed as well as humans did, but also whether the similarity of their communication strategy to the strategy humans use to communicate influenced their performance [36]. It was found that cooperation can be established without the use of theory of mind, but theory of mind can highly increase the efficiency of this process. The effectiveness of higher-order theory of mind is affected by the way agents choose which messages to send. In the right circumstances, higher-order ToM agents have the ability to choose messages that are less likely to be misinterpreted [36].

A notable difference between the agents in this simulation and humans is that agents are able to generate and compare a very large number of scenarios in a short amount of time before making their final move. This caused the second-order theory of mind agents to almost immediately reach a cooperation and solution. This is not a realistic goal for humans to achieve, regardless of how good they are at this game. This highlights some of the fundamental differences between ToM agents and humans.

2.1.3 Mixed-motive Settings: Coloured Trails

The third and final hypothesis for why higher-order reasoning may have evolved is the mixed-motive interaction hypothesis [30], which states that cognition and higher-order reasoning developed due to the need to perform in environments which are both competitive and cooperative, mixed-motive environments [28]. In such an environment, agents will work together in order to maximise their individual rewards. In order for agents to grasp the two-sidedness of such interactions and the types of deception that go along with it, they need higher-order theory of mind. This allows them to not just reason about true beliefs of other agents, but also about their false beliefs [34].

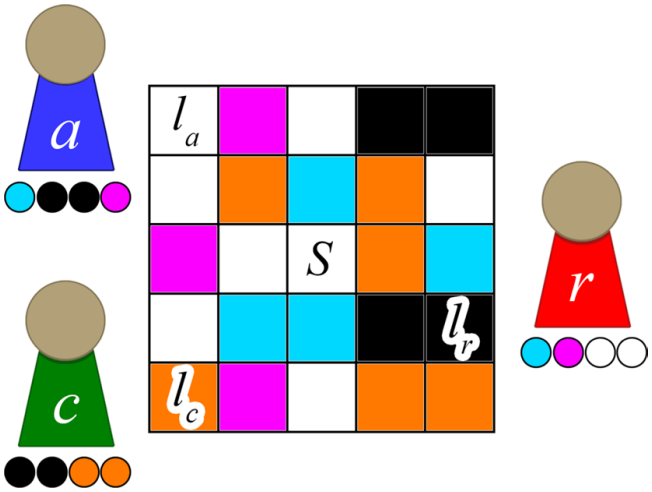


Figure 4: An example setup of a Coloured Trails game [34].

The effectiveness of theory of mind in a mixed-motive setting has been tested in several variations of the Coloured Trails game [14, 15]. In the Coloured Trails game, of which a visual example can be seen in Figure 4, agents need to traverse a board using coloured chips. Each tile in the board has a certain colour. Agents can move to tiles adjacent to their current tiles by "paying" with a chip that matches the colour of the tile they want to move to. The goal of the agents is to move from their starting location to their goal location in as few steps as possible. Since agents get handed a limited number of chips, they may need to trade to get the chip colours they need in order to get to their goal.

The first experiment uses one-shot negotiation in order to find out whether the allocator in a Coloured Trails setting benefits from using theory of mind, as well as to find out whether the use of theory of mind affects social welfare [34]. In this setting, it is especially useful to keep in mind that the ToM_0 agents used in this simulation were able to use associative learning, meaning that they were capable of improving and showing complex behaviour over the course of time. It was found that in this setting, theory of mind increases the benefits for both the allocator and the group as a whole. The benefit was largest for second-order theory of mind. No additional benefit was found for third-order Theory of Mind (ToM_3) [34].

In a second experiment, the effectiveness of theory of mind in a mixed-motive setting was tested in a variation of the Coloured Trails [14, 15] game in which two agents could negotiate through a series of offers [39]. This variation introduces a penalty for each additional round of negotiations, which encourages agents to reach an agreement faster instead of waiting for the "perfect" offer. Results showed that while it was possible for ToM_0 agents to reach a successful negotiation, they also had a tendency to "forget" about the cooperative aspect of the game, which often caused negotiation to fail [39]. ToM_1 agents performed better than ToM_0 agents, but not perfectly. They could prevent a failure of the negotiation, but still struggled not to let the competitive aspect of the game overtake the cooperative aspect. ToM_2 agents performed best in this setting, as they were a) able to balance the competitive and cooperative aspects without causing a failure of negotiation and b) able to switch between different orders of theory of mind based on their belief about the order of theory of mind of their opponent. This shows that higher-order theory of mind can be beneficial in mixed-motive settings with multiple rounds of negotiation.

A third experiment tested how humans interacted with different orders of ToM agents in a sequential variation of the Coloured Trails [14, 15] game, in an attempt to find out whether humans actually use higher-order theory of mind when playing this game, and to find out how successful these interactions can be [35]. Results showed that adults tend to use second-order theory of mind after only a few rounds, but that they do not often reach the Pareto optimal solution. The overall outcome was usually closer to the optimal solution when adults faced higher-order ToM agents than when they faced ToM_0 or ToM_1 agents. Additionally, adults showed behaviour consistent with second-order theory of mind when faced with higher-order ToM agents, even though they were unaware that their opponents would have varying levels of theory of mind.

A fourth experiment built on the previous ones by introducing unpredictability into the environment of the Coloured Trails game [37]. The aim of this experiment was to show whether using theory of mind is more effective in unpredictable negotiations than in predictable negotiations. This was done by looking at different combinations of static environments and goals together with dynamic environments and goals. The variation of Coloured Trails used for this experiment has the agents engage in one-shot negotiation. It was found that it is indeed advantageous to use higher-order theory of mind in unpredictable negotiation with ToM_1 and ToM_2 having the largest benefits. This advantage was attributed to the observation that ToM_0 agents struggle with unpredictability. This is likely also the reason that theory of mind is generally beneficial in competitive environments, as these environments encourage unpredictable behaviour.

2.1.4 Theory of Mind in Human-Machine Teams

While the scientific relevance of previous research on theory of mind cannot be denied, the work in this field has mainly been in the area of game theory and other well-defined problems. The aim of this project is to break out of the bounds of game theory and take theory of mind into the area of hybrid intelligence. The philosophy of hybrid intelligence is that the best parts of human and machine intelligence can be combined to reach new goals that have not been reachable by humans nor agents thus far [1]. Theory of mind can play a crucial part in this collaboration, as it provides agents with social skills that humans are accustomed to, making the communication and potential negotiation between these two parties more efficient and fruitful [1, 11].

In a first attempt to reach this goal, previous work has attempted to create a computational model of theory of mind abilities, with a focus on abstraction of beliefs and intentions [11]. The three-level abstraction procedure used for this model can be seen in Figure 5. Abstractions are meant to serve as a more efficient way of storing beliefs and intentions, but they are not meant as a replacement for the previous knowledge stored by agents. They serve as low-maintenance heuristics that are more easily accessible than using all previous knowledge, and are used by agents whenever possible.

Since previous research on theory of mind in hybrid systems is scarce, there are a lot of unknowns that can be discovered in the current project. It also means that a lot of what is known only applies to game-theoretic contexts. The aim of this project is to bring the use of theory of mind into the applied domain, specifically one which would profit from using hybrid intelligence. To account for the scarcity of previous research in the applied domain, the first step is to describe negotiation in energy regulation as a game. This will help create a computational model of agents negotiating using theory of mind in a hybrid system, in which humans and AI work together as a team.

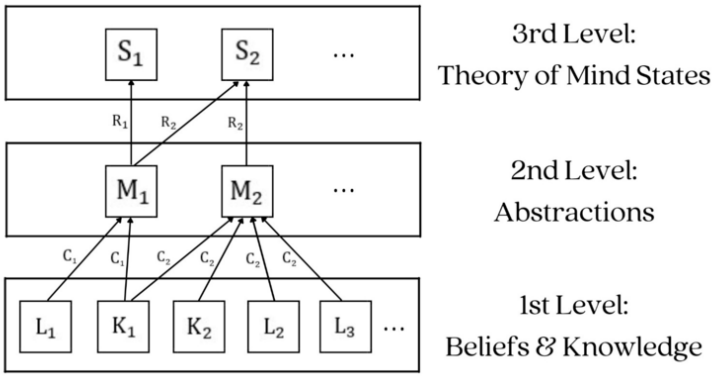


Figure 5: Theory of mind abstraction procedure [11].

2.2 Negotiation and Game Theory

In order to study the use of theory of mind in energy regulation using game theory, it is important to understand what negotiations are in the context of game theory. Negotiation is a mixed-motive setting in which parties want to come to an agreement even though they have conflicting goals and values [10, 12]. The parties both want to cooperate to reach an agreement, but they also want to compete to get the best deal possible. Besides having conflicting goals and values, other necessary features for a successful bargaining situation are that both parties feel that it is possible to reach an agreement and that they perceive that there are multiple ways to reach such an agreement [10].

Most negotiation follows a set structure [12], with some slight variation in the steps depending on the situation. The essential steps are as follows:

1. The presence of social conflict is recognised
2. Negotiation participants are identified
3. Private information is gathered and structured
4. Participants analyse their opponents
5. The negotiation protocol is selected and fixed
6. Iterated exchange of offers and counter offers

The following steps are optional:

7. Participants give argumentation for their offers
8. Participants learn from the information that is revealed by the offers of other participants
9. Participants alter their strategies as new information is revealed
10. An impasse occurs, all negotiation breaks down
11. Refinements are made to an agreed deal, often done to reach pareto optimality.

The parameters of negotiation are as follows [12]: the negotiation set, the negotiation protocol, the negotiation strategies, and the number of agents involved in the process. The negotiation set is the space of possible offers that can be made. The negotiation protocol is the set of rules that hold during

the negotiation. The negotiation strategy defines the proposals that participants make. The number of agents involved in the negotiation can take three forms: one-to-one, many-to-one, and many-to-many.

2.3 Energy Regulation

There are three layers of taxonomy in the world of grid-edge solutions [9]. In the first layer we consider agency, which considers whether units operate independently or not. There are two possibilities: direct control and indirect control. With direct control, there is a central entity that can access all information from all unite and that dictates their actions. With indirect control, units have autonomy at the local level over their actions.

Here, it is important to consider the following definition of "agents": "An agent is a computer system that is capable of independent action on behalf of its user or owner. In other words, an agent can figure out for itself what it needs to do in order to satisfy its design objectives, rather than having to be told explicitly what to do at any given moment. A multiagent system consists of a number of agents, which interact with one another, typically by exchanging messages through some computer network infrastructure." [42]. This means that if we consider agency in the context of multi-agent systems, we can expect there to be indirect control.

The second layer of the taxonomy considers how individual information is shared [9]. When considering direct control, the central unit knows everything about everyone so different ways of information sharing are not distinguished. When considering indirect control, there are three strategies: mediated, bilateral and implicit. In the mediated strategy, a central entity collects all information. In the bilateral strategy, units communicate information directly and bilaterally with each other. In the implicit strategy, there is no sharing of information among units. At most, units are informed of market information.

The third layer of the taxonomy [9] considers the game type: cooperative or competitive. In a cooperative game, units work together towards a central goal which is always prioritised over their individual objectives. In a competitive game, units only attempt to maximise their own utility.

Can Theory of Mind be used in negotiations about energy consumption performed by buildings? This was studied by creating a computational model of such negotiations based on game-theoretic principles.

Theory of Mind, the ability to reason about the unobservable mental content of others, is a promising concept in the world of multi-agent systems. The goal of this project was to test the potential of Theory of Mind in an applied setting, in this case urban energy sustainability, with the aim to prevent overloading the energy grid.

3 Methods

This chapter introduces the methods we used to answer our main research question. First, the negotiation game made to model the context explained in Chapter 1 is introduced, which we call Energy Trails. Next, the simulation of Energy Trails is outlined. Since the negotiation game is based on the game Coloured Trails [14, 15, 39], all sections also contain comparisons to Coloured Trails.

3.1 Introducing the Energy Trails Game

The first step towards creating a computational simulation of an environment in which agents represent buildings that negotiate about energy was to understand and simplify the problem. For this, we found a simple example context to make it easier to understand the problem we are attempting to solve. This context was introduced in Chapter 1 and is described in detail in the next section.

When creating a simplified model of the problem, we attempted to stay close to earlier research on theory of mind. Thus, we decided to adjust the Coloured Trails game as used and described by De Weerd, Verbrugge, and Verheij [39], which was based on the Coloured Trails game as described by Gal, Grosz, Kraus, Pfeffer, and Shieber [14, 15]. The game as used by De Weerd, Verbrugge, and Verheij [39] is described in Section 2.1.3.

3.1.1 Context

The example scenario that we chose is one close to home, on the Zernike Campus in Groningen[2]. From the buildings on the Zernike Campus, we chose two to be our example: the BB and the LB. Both of these buildings are managed by the University of Groningen, specifically by the Faculty of Science and Engineering.

We consider these two buildings specifically because they have very different needs and desires in terms of their energy consumption levels, especially considering how these are spread out over the day, even though they are very close to each other on the Zernike Campus.

The Bernoulliborg is the home of Mathematics, Artificial Intelligence and Computing Science. The building contains classrooms, computer labs, offices, a robotlab, and has an observatory on the roof. Most of the building is mainly used during the day, especially during the afternoon, but the observatory is exclusively used at night. The Linneausborg is the home of Biology, and contains many kinds of labs, classrooms, and offices. There are multiple labs in the Linneausborg that contain living animals and plants. This building is used during the day and during the night, and the temperature needs to be more or less constant, even during weekends, for the organisms that live in it.

Figure 6 shows the fictional time-based energy consumption profiles we created for these two buildings. The day is broken up into the morning, afternoon, evening and night. The desires of the agents are split up into different bins, which we will henceforth refer to as desire bins, which shows more clearly how the energy that they need is distributed over different use-cases. This will also make it easier to prioritise certain desire bins over others later. The desire-bin categories that we will focus on are: Heating, Ventilation and Air Conditioning (HVAC), Lights, Equipment, and Other.

[2]Zernikepark 1, 9747 AA Groningen

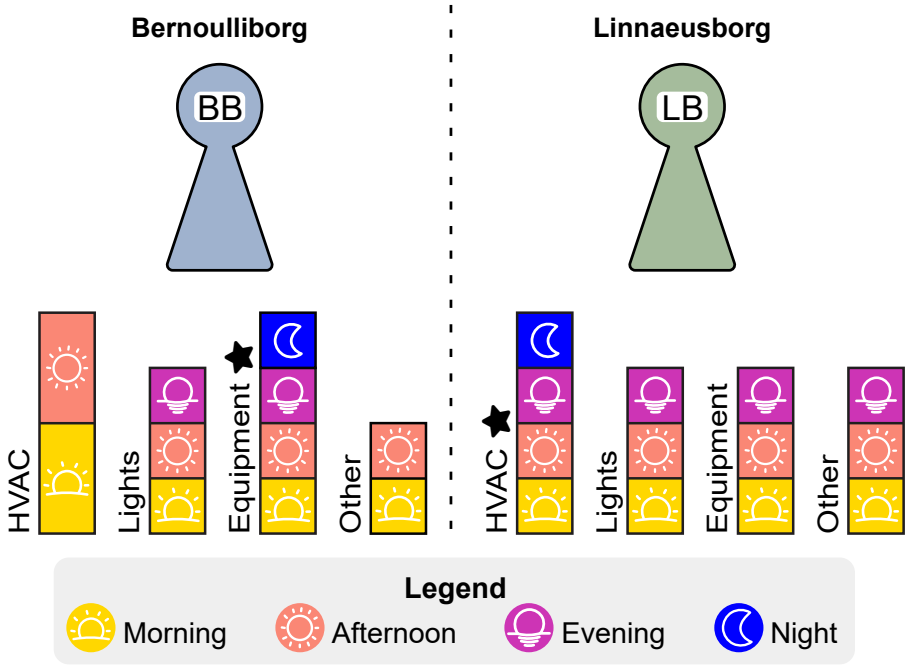


Figure 6: Example agents Bernoulliborg and Linnaeusborg and their desire bins. The use-case that is most important to each agent is indicated with ★.

3.1.2 Gameplay: From Coloured Trails to Energy Trails

This section will outline the general gameplay of our adjusted version of the Coloured Trails game and the ways in which it is different from the Coloured Trails game as used previously by De Weerd, Verbrugge, and Verheij [39]. The adjusted version of the game will henceforth be referred to as the Energy Trails game, or simply as Energy Trails. Explanations about how design choices relate to the energy domain can be found in Section 3.1.4.

The Players The Energy Trails game has n players which each represent a building in a network of buildings. Each of the players has desires that are distributed over different categorical desire bins, which hold the amount of energy desired and the time of day during which it is desired. The categories of the desire bins are the different use-cases that players plan to use their energy for.

An example of a set of players and their desire bins can be found in Figure 6, where the example players correspond to our context as explained in Section 3.1.1. Here, we see that agents desire to be able to consume certain amounts of energy during certain times of day for different use-cases. How much energy they need per time of day relates to how much area inside a bin is covered in one time-of0day's colour. The desire bins that these agents have apply to four specific use-cases: HVAC, lights, equipment, and other.

Since allowing the agents to use energy precisely as they desire would overload the energy grid during certain times-of-day, we need to find a way to let them negotiate about their desired energy consumption levels per use-case, with a focus on the time-of-day during which they wish to consume this energy. This is why we used these desired amounts of energy consumption as the goal in the Energy Trails game, which we call "goal compositions", rather than the "goal locations" that we saw in Coloured Trails. Each individual goal composition is derived from a different desire bin. The complete collection of goal compositions that an agent aims to reach is called a "goal profile".

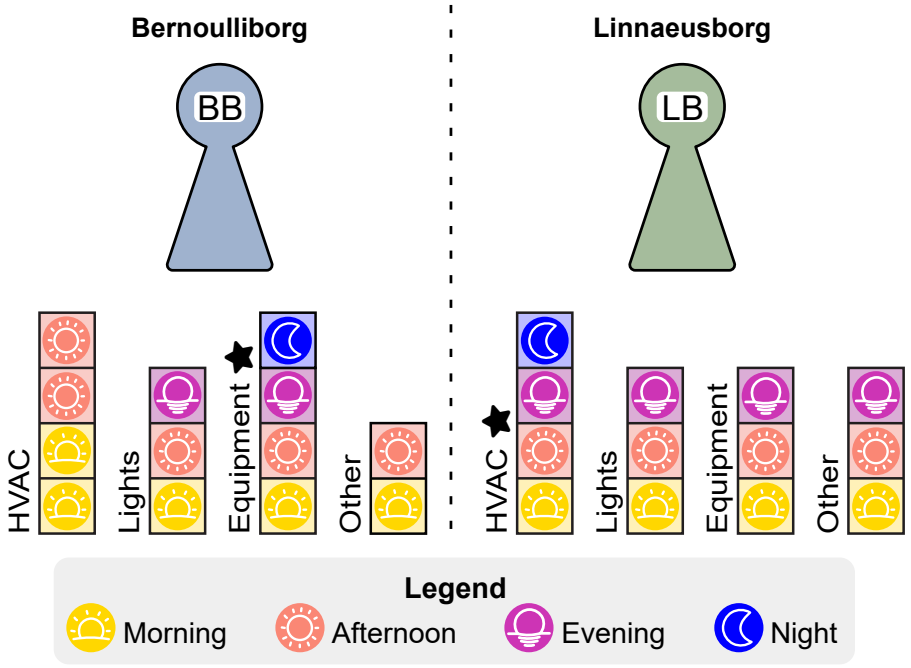


Figure 7: A visualisation of how the individual desire bins of each agent are translated into a goal composition per use-case. The use-case that is most important to each agent is indicated with ★.

The way the individual desire bins are translated into goal compositions can be seen in Figure 7, which finally turns into one goal profile per player as seen in Figure 8. The desires are split up into equal parts that roughly translate to one chip's energy "value". When using this method on real data, the chips need to be assigned a value through which the bins can be split up into chips. For the example outlined in section Section 3.1.1, the chips will have no corresponding real "value" as the desire bins of the BB and LB are not based on real data. If it is important for the agents to have a chance to find an optimal solution in a game, the total number of chips in the desires of both agents should not be higher than the total number of chips in the game.

In Coloured Trails, a player aims to reach its goal location. In Energy Trails, a player wants to make as many trails that reach the goal compositions in their goal profile as possible, in order to be able to use the desired amounts of energy during certain times of day for certain applications. An example of such a goal composition and the way a player can reach their goal composition on a board can be seen in Figure 9. In this case, if the agent in the example reaches their goal this means that they can use one unit of energy each during the morning, afternoon, evening, and night for equipment. An agent gets points for creating a trail with at least its goal composition, in which the order of the chips does not matter.

The goal composition of an agent's trail for a board is based on its corresponding desire bin. Players only benefit from reaching the goal composition on a board that matches the use-case of that goal composition. For example, the "Equipment" goal composition should only be reached on the "Equipment" board.

The Boards Each individual board used in Energy Trails is similar to the board used in Coloured Trails. Each board is a grid of $m \times m$ tiles with o different colours. The colours of the chips and tiles in the Energy Trails game represent the times of day that are also present in the players' desire bins, as explained above.

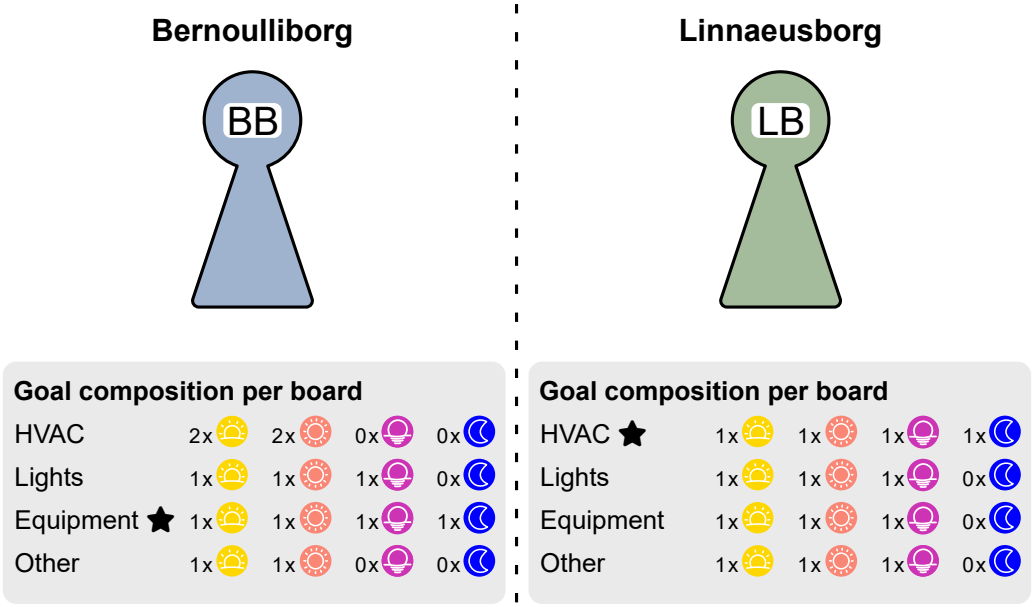


Figure 8: The goal compositions per player, based on the bins of the players in the Bernoulliborg and Linnaeusborg example. Each individual box labelled "Goal composition per board" represents a goal profile. The main goal composition, which is prioritised over the other goal compositions, is indicated with ★.

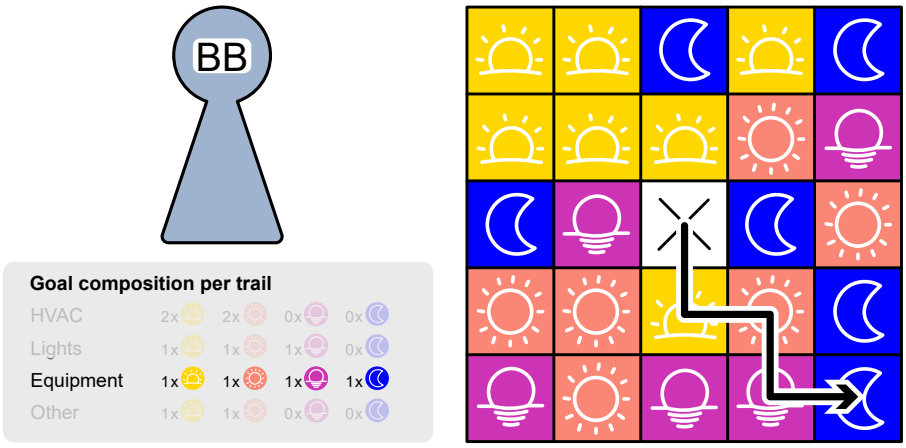


Figure 9: An example of how a player can fulfil one of their goal compositions on an example board.

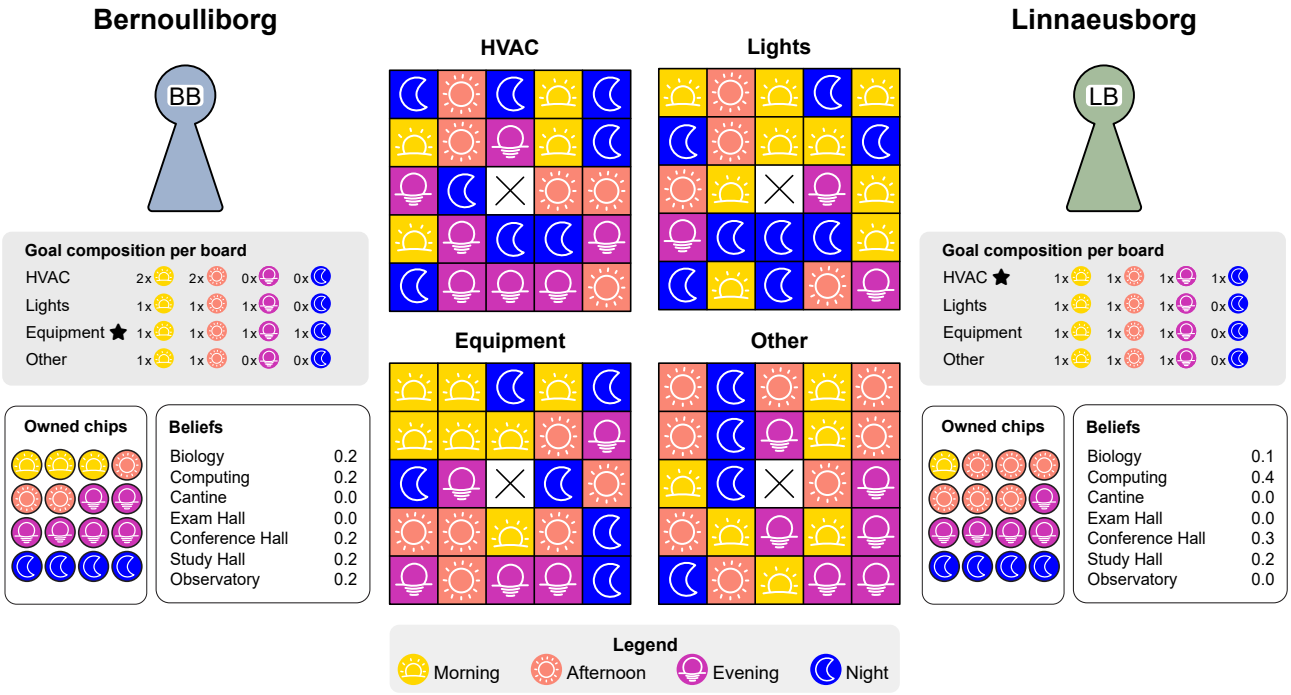


Figure 10: The Energy Trails game with parameters based on the Bernoulliborg and Linnaeusborg example: 2 players ($n = 2$) and 4 boards ($p = 4$). The chip and tile distributions and placement in this example were generated randomly.

Unlike the Coloured Trails game, which only has one board, the Energy trails game has p boards, where $p > 1$. Each board is associated with a use-case. The use-cases that the boards represent match the use-cases that the goal compositions within each player's goal profile represent. Overall, the use-cases need to be the same across the goal profiles of players and the boards. Separating the boards to match the use-cases allows players to differentiate and prioritise between the use-cases they may want to allocate energy to.

An example of players, as described in Section 3.1.1, with their desire bins can be seen in Figure 6. The corresponding energy profiles of these players can be seen in Figure 8. The game that these agents play would have four boards, which correspond to the four different use-cases represented in the desire bins and goal compositions of the agents: "Heating, Ventilation and Air Conditioning (HVAC)", "lights", "equipment", and "other". An example of a setting with these four boards and our two example players, Bernoulliborg and Linnaeusborg, can be seen in Figure 10.

By having p boards that correspond to the use-cases of the goal compositions of agents, we are able to prioritise one of the goal compositions in their goal profile over the others. Completing the main goal composition, which will henceforth be referred to as "reaching a main goal", gives players a higher number of points than completing a regular goal composition, which will henceforth be referred to as "reaching a regular goal". In Figure 6, 7, 8, and 10, the main goal desire bins and main goal compositions respectively are indicated with the symbol ★.

At the start of the game, the chips need to be generated, the chips need to be distributed over the players, the tiles need to be generated, and the tiles need to be distributed over the boards. The number of chips available of each colour does not need to be equal to the number of tiles of each colour. The way distributions of the tiles and chips are decided depends on the real-life situation that

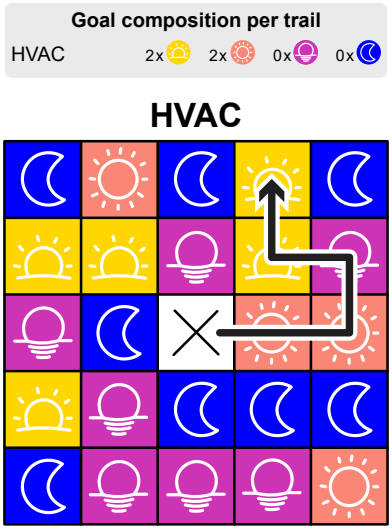


Figure 11: Example of a situation in Energy Trails in which the goal composition cannot directly be translated into a trail on the board, forcing a player to create a trail that contains more chips than their goal composition.

the game is attempting to model, if any.

There are q chips in the game, where $q = n \times p \times (m-1)$ to allow all players to create a path that, in theory, reaches the "end" of each of the boards. Whether this is actually possible depends on the goals of agents, the tiles on the boards, and the chips that agents have. The chips and how they are distributed over the players are what players negotiate about in the game.

When the tiles are distributed over the boards, this creates a patchwork of different colours where every location on the board can have one coloured tile. The chips can be used to traverse the board. For a player to "walk" a path over a board, they need to have a chip in the same colour as each tile they intend to step on. Each step on the board can only be horizontal or vertical.

Depending on how the tile placement is generated, it may not be possible for an agent to walk a trail exactly matching its goal composition on a board, even if it has the right chips. An example of this can be seen in Figure 11, where we see that there is no trail on the board that corresponds exactly with the goal composition for that board. In these cases, an agent is allowed to take additional steps; as many are necessary to walk the shortest trail on the board that contains its full goal composition. In this specific example we see that the Bernoulliborg cannot reach its HVAC goal composition without using any chips that are not in the goal composition. In this case, it can only reach its goal composition by taking a small "detour" of one evening chip. How this relates to situations in the energy domain is explained in Section 3.1.4.

The Negotiations Negotiations in the Energy Trails game are almost the same as in the Coloured Trails game. Players participate in negotiation rounds to decide on the final distribution of the chips in the game over the players. They negotiate about all chips in the game at once, which means that each offer contains a suggestion for which player each chip in the game should belong to. An example of what an offer looks like in the Energy Trails game can be seen in Figure 12.

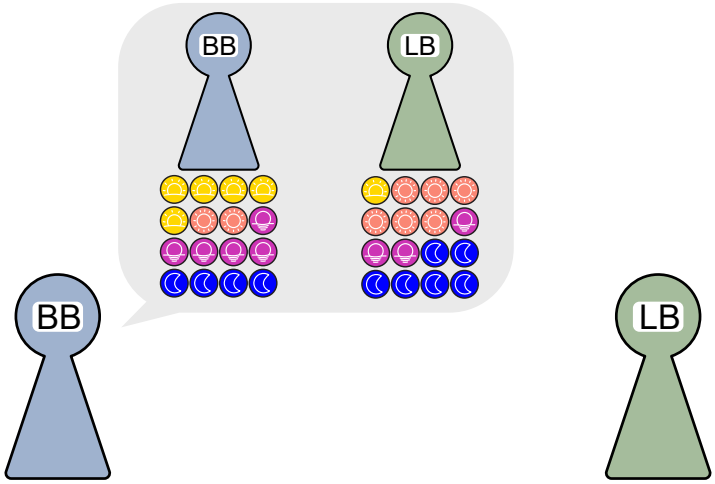


Figure 12: Example of an offer placed by player

The Energy Trails game consists of multiple rounds of negotiations. Traversing the board only happens after the negotiation is finished in order to calculate the points of each player. The starting player places an opening offer. After that, agents alternate between choosing one of three actions until a player ends the negotiation. The actions from which the players can choose are: 1) Accept: a player accepts the previous offer which ends the negotiation and redistributes the chips according to the offer that was accepted. 2) Reject/withdraw: a player rejects the previous offer and withdraws from the negotiation, which ends the negotiation and leaves the chip distribution as it was at the start of the game. 3) Counteroffer: a player places a counteroffer, the negotiation continues. When the negotiation ends, the chips are distributed (if applicable), agents walk their chosen paths, based on those paths the scores are calculated, and the game ends.

The information in this game is not complete. Agents can observe the chips owned by the other player and the tile distribution on each board, but they cannot observe the goal profiles, nor the goal compositions, of other agents. This is what they attempt to figure out using theory of mind. To illustrate, Figure 13 shows what an example game would look like from the perspective of the Bernoulliborg. It can see its own chips, the chips owned by the other player, and the boards. It can not see the goal profile, nor the goal compositions inside it, of its trading partner, in this case the Linnaeusborg. The assumption is also made that the agents cannot see the desire bins from which the goal compositions are derived, since this would allow them to deduce what the goal profile of their trading partner is.

Figure 10 also shows the beliefs of the players. These beliefs are a representation of the theory of mind-based strategy that the example players are using. While this is representative of the experimental set-up we will be using and of how players will approach the game in our simulation, it is entirely possible that players of the Energy Trails game employ a different strategy. The simulation and the beliefs will be explained in detail in Section 3.2.

3.1.3 Scoring

This section explains how the scoring works in Energy Trails. The reasoning behind the scoring, particularly the prioritisation of each event that can occur in the game, is explained in this section. A list of the scoring in order of priority per variation can be found in Table 2.

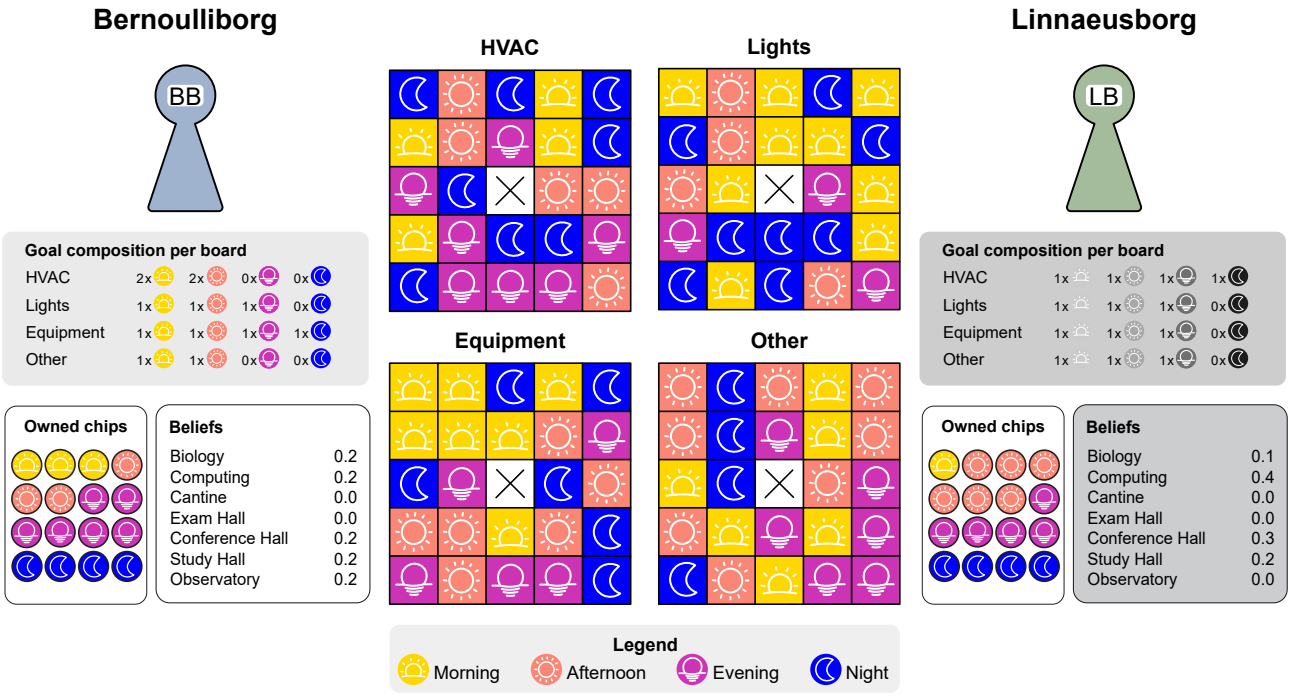


Figure 13: An example of a game played by Bernoulliborg and Linnaeusborg. The grey areas are not visible to the Bernoulliborg.

Points are only awarded for reaching goal compositions, not for reaching all compositions in a goal profile. Thus, in this section the word "goal" always refers to a goal composition within a player's goal profile. "Main goal" refers to the main goal composition in a player's goal profile.

The two different variations of Energy Trails are distinguished through the scoring: a competitive variation and a cooperative variation. What we mean by this, is that the scoring decides whether the focus of the game is on being cooperative or on being competitive. Overall, the game is still mixed-motive, regardless of the scoring used.

The competitive scoring is nearly exactly the same as the scoring used in the version of Coloured Trails used by De Weerd, Verbrugge, and Verheij [39], with the exception of the points awarded to a player for reaching their own main goal. This is due to main goals being a new feature in the Energy Trails game that was not present in the Coloured Trails game.

Since the original Coloured Trails did not differentiate between main goals and regular goals, all goals awarded the same number of points. To ensure that players prioritise their main goals in the Energy Trails game, reaching a main goal is awarded 1000 points. This is more than reaching your own regular goal, which is awarded 500 points.

The cooperative scoring is also based on the scoring used in the version of Coloured Trails used by De Weerd, Verbrugge, and Verheij [39], but was extended to feature point sharing between players, as well as increased scoring for reaching a main goal.

In order to motivate players to prioritise the main goals of other players, they receive a higher reward when their trading partner reaches their main goal than when they reach their own regular goal. Since they should still prioritise their own main goal over that of their trading partner reaching their main goal, the value of this reward is placed in between the reward they get for reaching their own main

Event	Competitive	Cooperative
Reaching your own main goal	+1000	+1000
Another player reaching their main goal	+0	+750
Reaching your own (regular) goal	+500	+500
Another playing reaching their own (regular) goal	+0	+250
Step towards your own goal	+100	+100
Leftover chip	+50	+50
Another player reaching their goal using a purchased chip	0	0
Another playing taking a step towards their goal	0	0

Table 1: The points system used in Energy Trails.

Competitive		**Cooperative**	
1. Reaching your own main goal	1000	1. Reaching your own main goal	1000
2. Reaching your own regular goal	500	2. Trading partner reaching their main goal	750
3. Step towards your own goal	100	3. Reaching your own regular goal	500
4. Leftover chip	50	4. Trading partner reaching their regular goal	250
5. Trading partner reaching their main goal	0	5. Step towards your own goal	100
6. Trading partner reaching their regular goal	0	6. Leftover chip	50
7. Trading partner step towards their goal	0	7. Trading partner step towards their goal	0

Table 2: Prioritised score list for each variation of Energy Trails.

goal, and the reward for reaching their own regular goal.

Similarly, players are encouraged to prioritise reaching their own regular goal over their trading partner reaching their regular goal. In the cooperative version, a trading partner reaching their regular goal is placed such that it is below a player reaching their own regular goal, but above taking a step towards your own goal. This way, players will only be selfish with their chips if neither player is close to reaching any of their goals.

No points are awarded to a player when their trading partner takes a step towards their goal. This was done to allow players some sense of competitiveness in their gameplay, since this will make them prefer having their own leftover chips over letting the other player take steps towards their goal. Leftover chips are placed below taking a step towards your own goal, as is the case in Coloured Trails.

3.1.4 Relation to the Energy Domain

This section outlines how the design choices made in the Energy Trails game relate to the energy domain in terms of energy regulation with regard to a group of buildings. The comparisons made with real-word circumstances were retrieved from conversations with domain experts during the time of this research.

In the real world, energy regulation within a group of buildings that are close to each other can be quite cooperative [9], which goes against the nature of most games. Since all buildings rely on the same grid for their energy, they rely on each other's choices for their own energy consumption. On top of that, bad conditions in a neighbouring building may reflect badly on your own. This means

that the people who manage the energy consumption levels in those buildings do need to consider the energy consumption of their neighbouring buildings.

This relates to why Energy Trails is necessary in the first place: it creates a mixed-motive setting in which the buildings can negotiate about how much energy they can use and when, whilst trying to get the highest score for themselves. Next to that, one of the reasons we introduced the cooperative variation of Energy Trails was to test how well translating these circumstances into the game directly works in the system.

Another reason the cooperative variation of Energy Trails was introduced is related to efficient division of resources. Claiming more energy than necessary has different consequences depending on the scale that is considered. On a large scale, it can lead to very large fines. On a small scale, for example in households, there are no immediate consequences. On the scale of the Bernoulliborg and Linnaeusborg example that we are considering, there should be an incentive not to go too far over, but there are no extreme consequences if it does happen. This is represented in the cooperative scoring. In the cooperative scoring, another player reaching a goal is prioritised over having leftover chips. This encourages players of the game to consider the goals of the other agents before claiming all chips themselves.

In order to find a compromise between buildings, it may be necessary to have the option to prioritise certain goals over others. This ensures that the main usage of a building is not jeopardised. This is why the main goals were introduced. That way, buildings have a way to differentiate between goals they cannot afford to compromise on and goals that have lower importance.

The way the distribution of chips and tiles on the board is generated can also relate to real-world circumstances. In general, the chip distribution, meaning the number of chips available for each time-of-day, should reflect the availability of resources. As explained before, the goal compositions of agents reflect their energy desire bins. The tile distribution on the boards can represent external constraints that may influence how much energy a building needs to use, such as the weather, large events, or contractual obligations.

The way humans can interact with the system is by deciding which chips and tiles are available in the game and by distributing them. Humans can create the chip and tile distributions of the system to correspond to the real-world circumstances. This way, they can exert influence on the system without directly interacting with the negotiation. The buildings' desire bins can also be crafted by a human, but should ideally be derived from scheduling and historic data.

Observability in the game is also influenced by circumstances we find in the energy domain. Companies and institutes are often hesitant to share their data and goals with others in order to protect their privacy. This can have many reasons, but can also just be for a sense of security. This is why the goals of other agents are not observable and are what is reasoned about using Theory of Mind.

3.2 Implementation

This section will outline how the computational simulation of Energy Trails works, which can be found on GitHub[3]. This simulation of Energy Trails is based on the existing simulation of Coloured Trails by De Weerd [34], which will also be outlined in this section to provide a basis for understanding the simulation of Energy Trails.

[3] https://github.com/I-Tilleman/MSc-EnergyTrails

Overall, the simulation contains roughly three components:

1. The Negotiation
2. The Game
3. The Agents

The Negotiation handles the gameplay itself and calls on the agents to take actions in the game. The Game handles the creation of the game environment by setting everything up and distributing resources over the players. The Agents handle offers, counteroffers, and belief updates and take actions and update their internal states accordingly. The Negotiation is an integral part of this simulation because without it, neither The Game nor The Agents have meaning. The Negotiation and The Game together act as a simulation of Energy Trails. The Negotiation and the Agents together act as a simulation of players that play Energy Trails.

First, we will take a closer look at the simulation of Coloured Trails and the implementation of ToM agents in this simulation. This will help with understanding the implementation of Energy Trails, as the mechanics are similar, but Energy Trails is more complex.

3.2.1 Coloured Trails

Before we look at the implementation of Energy Trails and at how the ToM agents work in Energy Trails, it is important to understand how they worked in Coloured Trails. The mechanics of the negotiations and belief updates in negotiation agents have not changed much in Energy Trails compared to Coloured Trails, but since the scale has increased drastically, it is easier to grasp the concepts when they are explained in Coloured Trails first.

This section is based on the Python implementation of Coloured Trails by De Weerd [34]. Instead of "seeing" a board with locations, agents in the simulation of Coloured Trails see a list of possible goal locations that the other player can have, as well as a list of offers that they can make and their utilities. They can also see the set of chips owned by their trading partner. The list of possible goal locations that an agent can have does not contain all locations on the board, but rather the ones visualised in Figure 14. These locations are all linked to their own full utility function in which agents can see which offers are good for which goal locations, which is also how it works for the goal profiles seen in Figure 18 for Energy Trails.

In general, all agents want to find a way to achieve a division of chips that maximises their score. All agents share a basic mechanism for how to judge which offers to place and accept. In general, agents will only consider placing offers that increase their score compared to the score they would get from the chips that they start the game with. If agents do not consider it possible to place or receive an offer that increases their score, they will withdraw from the negotiation. Beyond that, the way agents choose which offers to place or accept depends on the order of theory of mind that the agents use.

A ToM_0 agent uses its own list of offers sorted by utility, as well as its previous experience of which kinds of offers are accepted. The initial offer made by a ToM_0 agent, before it has had time to learn from experience, is usually an offer in which the agent making the offer gets almost all chips, and the trading partner gets almost nothing. This is because that is the offer which has the highest utility for the agent. Since this agent does not have a concept of its trading partner having a goal location, it can only know what to offer based on which previous offers were rejected or accepted. It has some basic ability to generalise this, as can be seen in Example 1. How strongly it learns from previous

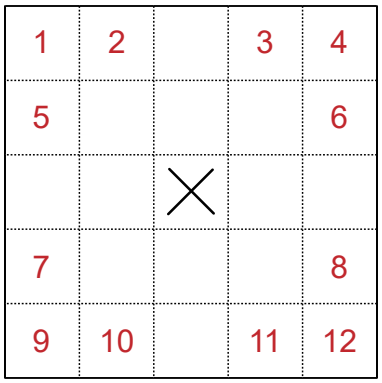
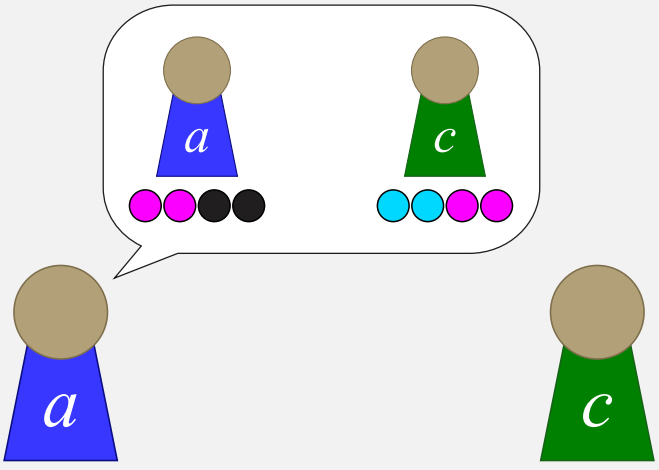


Figure 14: The possible goal locations on the playing board of Coloured Trails.

experience is affected by the *learning rate*. If the learning rate is too low, it will not learn from experience and will simply go through its offer list in the order of highest to lowest utility until an offer is accepted.

The way a ToM_0 agent learns from experience is by considering the chip difference between an offer and the starting chip division. This way, it can learn which chips are preferred in offers, as well as what number of chips is preferred. Using what it has learned, it forms beliefs about whether it expects its trading partner to accept each potential offer.

Example 1: Zero-order Theory of Mind in Coloured Trails
Let us consider the situation illustrated below. In this illustrative example, agent a is a ToM_0 agent. If agent c declines this offer made by agent a, agent a knows that offers in which it offers agent b equal to or less than $\{blue, blue, magenta, magenta\}$ will not be accepted. Because a ToM_0 agent is able to generalise this type of knowledge, it now also knows not to place the following offers where player b strictly gets: $\{blue, magenta, magenta\}$, $\{blue, blue, magenta\}$, $\{magenta, magenta\}$, $\{blue, blue\}$, $\{magenta, blue\}$, $\{blue\}$, $\{magenta\}$. It knows this because these offers are strict subsets of the original offer. This means these offers will have a lower value to the trading partner than the offer it already placed, which got rejected.



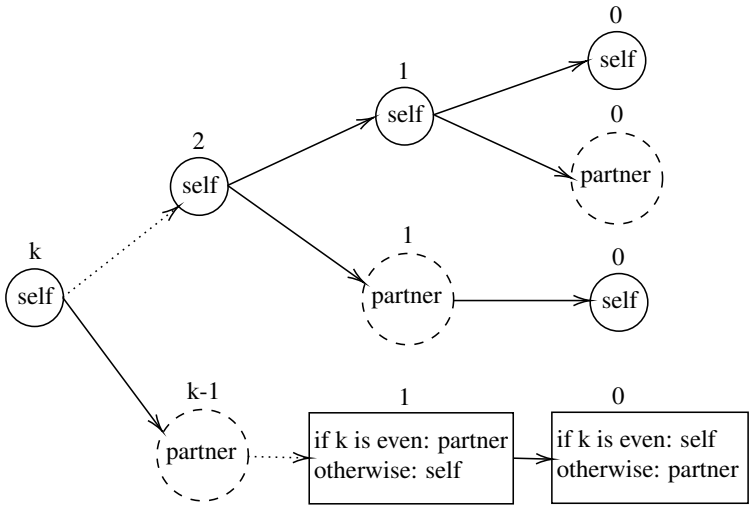


Figure 15: Model structure of a ToM_k agent. A ToM_k agent, $k \geq 1$, models its trading partner as a ToM_{k-1} agent. A partner model is always confidence-locked, which means that a modelled partner does not have a (direct) model of itself with a lower-order theory of mind capability. An agent that is not confidence-locked also has a model of itself where it can use its lower-order theory of mind capability. Note that the kth-order beliefs of an agent with at least a kth-order theory of mind capability of whether an offer is going to be accepted are the modelled $(k-1)$th-order beliefs of the trading partner (where $k \geq 1$) [7].

ToM_1 agents and ToM_2 agents do not strictly base their next offers on which kinds of offers were accepted or rejected in the past, but rather use their beliefs about the goal location of their trading partner because they have an inherent awareness that their trading partner has a goal location. These beliefs are updates based on which offers were placed in the current negotiation. This awareness that their trading partner has a goal is used to attempt to figure out the goal location of their trading partner. This knowledge is used to make offers in the future, in order to not only optimise their own score, but to also ensure that their offer is accepted.

ToM_1 and ToM_2 agents use their beliefs about the goal locations of their trading partners by running a full simulation of the offer selection procedure of their trading partner when they calculate the value they want to assign to each possible offer based on their current beliefs. In order to do this, these ToM agents need to have internal simulations of their opponents as well as of themselves. This means that each k-th other Theory of Mind (ToM_k) agent with $k \geq 1$ has internal simulations of themselves and their trading partners as if they were a ToM_{k-1} agent. The same holds for those simulated agents, as long as $(k-1) \geq 1$, except that they will not have a simulated version of themselves, since they are confidence-locked. An example of this can be seen in Figure 15. Bear in mind that in Energy Trails, the maximum order of theory of mind used by agents is two.

ToM_1 agents try to figure out the goal location of their trading partner, which is represented by a utility function. Then, they use an estimation of their trading partner's goal location and take this into account when making an offer, based on the idea that their trading partner is a ToM_0 agent and thus bases its offers mostly on its own utility.

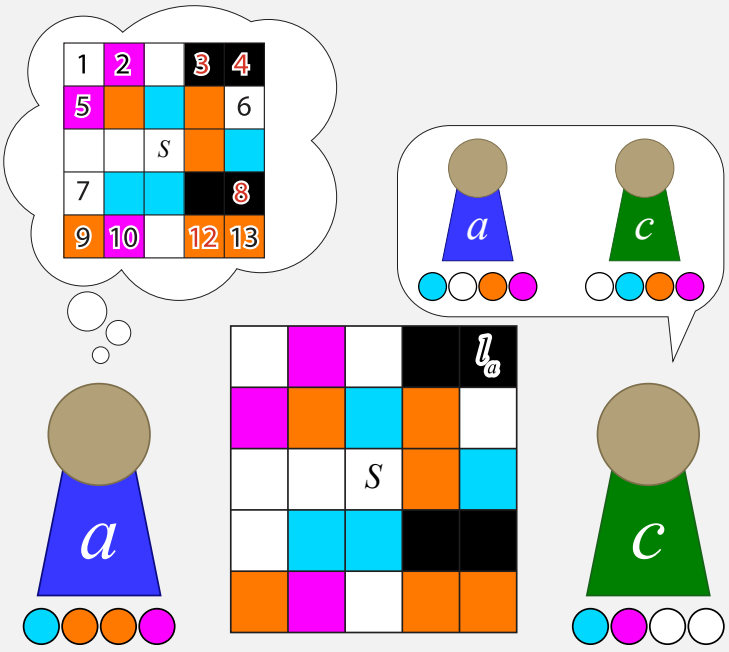
The initial offer made by a ToM_1 agent is also based on its estimation of its trading partner's goal location. Since this estimation is based on offers made by the agent's trading partner, and there may not have been any yet, the estimation is an average of all possible goal locations that are possible in

the game. This means its offer will be based on the trading partner having the average of the goal locations as its goal location.

When the trading partner makes an offer, the ToM_1 agent uses this offer to deduce which utility function its trading partner might have. This means that when the ToM_1 agent makes its next offer, the list of possible utility functions that its trading partner could have will have been reduced, making the estimation more accurate. An example of this can be seen in Example 2.

A ToM_2 agent does not only know that its trading partner has a goal location, but also that it could be a ToM_1 agent trying to figure out the utility function of the ToM_2 agent. It uses this information to try to make offers that reveal its own utility function as much as possible, such that its trading partner is able to make an offer that is good for both agents. This is done by simulating how a ToM_1 agent would respond to its offers. Since the ToM_1 agent will assume that the ToM_2 agent is actually a ToM_0 agent, it will assume that the ToM_2 makes an offer that is optimal for itself to reach its own goal location and will then adjust its beliefs about the utility function of its trading partner accordingly.

Example 2: First-order Theory of Mind in Coloured Trails
Let us consider the situation illustrated below. Agent a is a ToM_1 agent, which means that it is trying to figure out the goal location of agent c, l_c.
Upon receiving the offer as seen in the image below from agent c, it draws the following conclusions about the goal location of agent c: the locations 3, 4, 8, and 12 are unlikely to be the goal location of agent c. This means that, assuming that it still considered all goal locations to be equally possible before this offer was placed, the number of goal locations agent a considers possible for agent c has been reduced from 12 to 8.

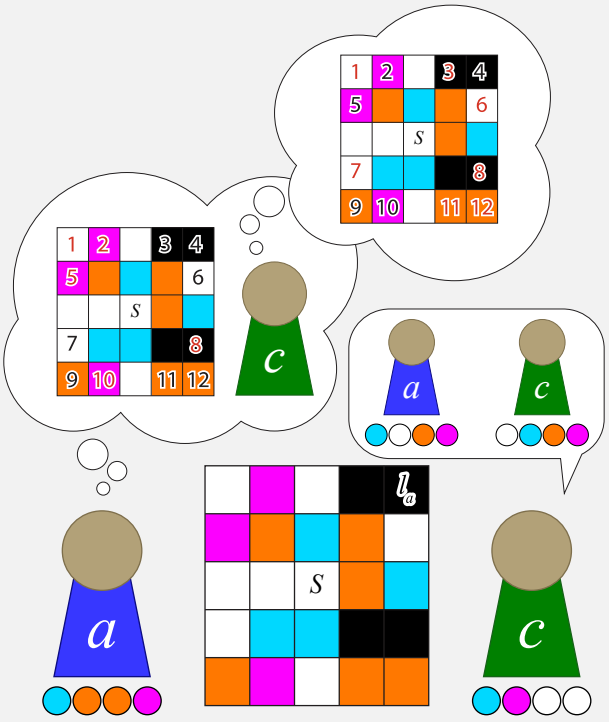


When the ToM_2 agent makes its initial offer, it tries to reveal information about its own goal location. Once it has received a counteroffer, it can deduce whether it needs to be clearer about its own utility function and can deduce information about the utility function of its trading partner. In making future offers, it can use this information to optimise the outcome further. An example of how a ToM_2 agent uses information from previous offers to place new offers can be found in Example 3.

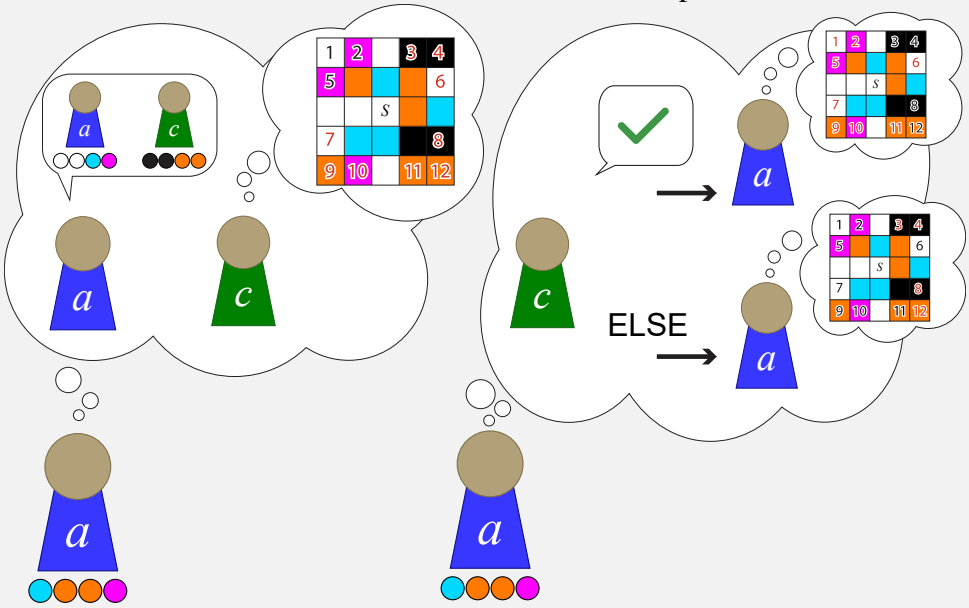
Example 3: Second-order Theory of Mind in Coloured Trails

Let us consider the situation illustrated below. Player a is a ToM$_2$ agent who is trying to figure out the goal location of player c, l_c, as well as if player c has figured out its own goal location, l_a, yet.

Upon receiving the offer from player c, player a draws the following conclusions: l_c is likely to be 3, 4, 6, 7, 9, 11, or 12, and player c probably does not know l_a yet, but rather thinks its 2, 5, 4, 9, or 10.



In order to inform player c of its goal location, and to try to give a good offer based on the location knowledge it currently has of l_c, player a places its offer. The chips it allocates to itself will exclude 4, 9, 10 from what player a expects to be player c's internal list of options for l_a. It also adds location 1 as an option.



The chips it allocates to player c are a good offer in case l_c is 3, 4, 8, 12. If this offer is accepted it will know that l_c was likely one of those locations, if not it knows that l_c is unlikely to be one of those locations. Thus, it can exclude a number of options for l_c based on player c's reaction to this offer.

Generally speaking, this means that theory of mind is used at two points in the process. When an offer comes in, theory of mind is used to interpret what the other player is trying to do. When making an offer, theory of mind is used to predict how the other player will respond.

3.2.2 Energy Trails

This section describes the simulation of Energy Trails that was created for this thesis based on the simulation of Coloured Trails by De Weerd [34] as described in the section above.

The Parameters First, we will look at the parameters used for this simulation. Whereas some of them are in line with the Coloured Trails game as used by De Weerd, Verbrugge, and Verheij [39], others were chosen to fit the Bernoulliborg and Linnaeusborg example as described in Section 3.1.1. A representation of the game with the parameters as outlined in this section can be found in Figure 10.

Each board is a grid of 5×5 tiles with four different tile colours, corresponding to four times of day: morning, afternoon, evening, and night. This way, agents can create routes that contain energy chips corresponding to all different times of day. There are four boards, which correspond to the use-cases of the players' desire bins: HVAC, Lights, Equipment, and Other.

This game contains 32 chips. At the start of the game, the two players get 16 chips each. The distribution of chips in the game, the number of chips available per time-of-day, is done randomly. The distribution of those chips over the players is also done randomly. Similarly, the distribution of tiles, the number of tiles for each time-of-day, and the distribution of those tiles over the boards is done randomly.

While it would have been possible to base the distribution of chips and tiles on real data about energy availability during different times-of-day and external factors that affect how much energy is consumed, this would have restricted the number of negotiations in which we could test our ToM based on the amount of available data. Thus, we decided to go for random generation instead.

In the implementation of Coloured Trails, there is a set number of goal locations. Hence, we decided to do the same for the goal profiles. Here, there was a choice to be made. One possibility was to have a set of building profiles with set goal compositions and main goal compositions, of which an example is given in Figure 16. Another possibility was to have a set of goal compositions of which all combinations would be considered possible for the possible goal profiles. An example of this can be seen in Figure 17. This would require agents to either reason about the individual goal compositions per board, which is difficult because agents only 'use' chips after the negotiation ends, so agents do not know how their trading partner will distribute chips over boards, or about all possible combinations of the building profiles, which would lead to a very large search space.

Thus, the choice was made to go with the first option, and to create a few goal profiles with set goal compositions and set main goal compositions in Energy Trails. Agents do not reason about main goal compositions separately, as they are part of the goal profiles. The goal profiles used in the simulation,

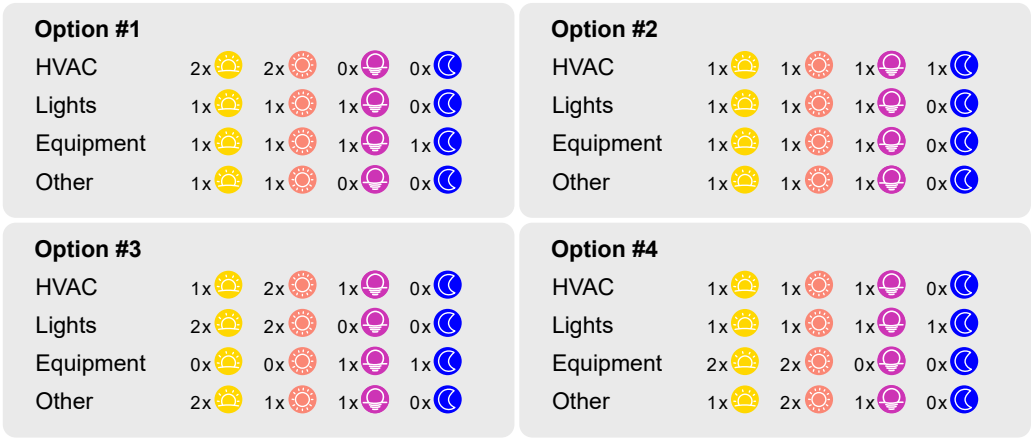


Figure 16: An example of the possible goal profiles if the possibilities consider all boards as one group.

as seen in Figure 18, diverge slightly from the context as described in Section 3.1.1, but still fit within the theme of buildings on a campus.

Just like in the work by De Weerd, Verbrugge, and Verheij [39], a negotiation limit was implemented. This means there was a limit to the number of negotiation rounds that agents could use within one negotiation. In earlier work, this limit was set to 100. In the simulation of Energy Trails, we set this limit to 25. The limit was decreased due to time constraints. The number 25 was chosen because it was found that in situations where agents frequently reached the original negotiation limit of 100 rounds, they would never conclude the negotiation after reaching at least 25 negotiation rounds.

The simulation contains an additional parameter that is not directly related to Energy Trails, but rather specific to the negotiation agents: the learning speed. The learning speed determines how heavily new information affects the beliefs of an agent. This is set to 0.8, which is the same as in the work by Weerd, Verbrugge, and Verheij [39].

The board size, number of chips per agent per board, the competitive scoring (except for points awarded for reaching the main goal), and the learning speed are all the same as in the Coloured Trails game as used by De Weerd, Verbrugge, and Verheij [39]. The number of boards, the number of tile colours, the total number of chips, the cooperative scoring, and the goal profiles were chosen to fit the Bernoulliborg and Linnaeusborg example as described in Section 3.1.1 as well as the energy domain in general.

The Negotiation This component of the simulation handles the gameplay itself: initialising the agents and game, making the agents negotiate, keeping track of the number of negotiation rounds, and stopping the negotiation when the negotiation round limit is reached. As the game and agent classes are initialised here, the parameters for those classes are also set, including: agent ToM levels, backup frequency, negotiations, round limit, number of boards, and whether the competitive or cooperative scoring is used.

During the negotiation, agents can take one of three actions: 1) Accept, which ends the game with the accepted offer being the new chip distribution. 2) Reject/Withdraw, which ends the game with the starting chip distribution being the new chip distribution. 3) Counteroffer, where an agent plays another offer, this continues the game. The execution of these actions is done by the agents themselves,

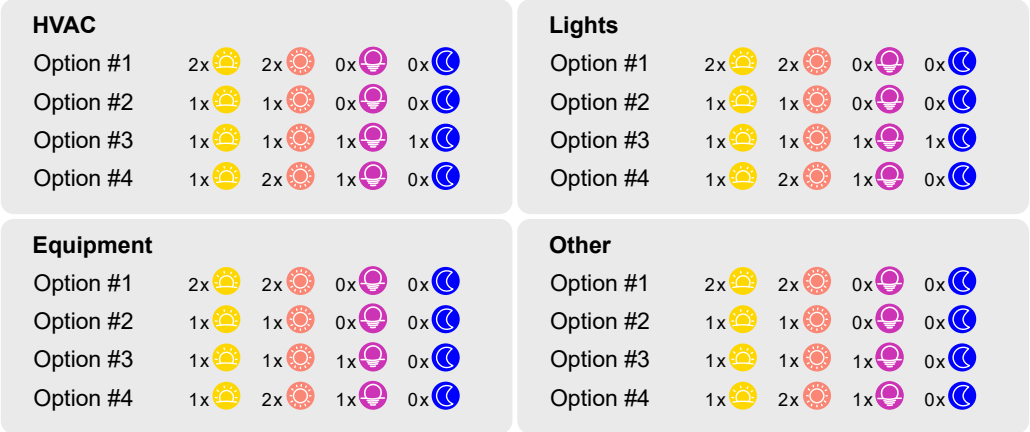


Figure 17: An example of the possible goal compositions, grouped by use-case rather than by goal profile, if the possibilities consider all boards individually.

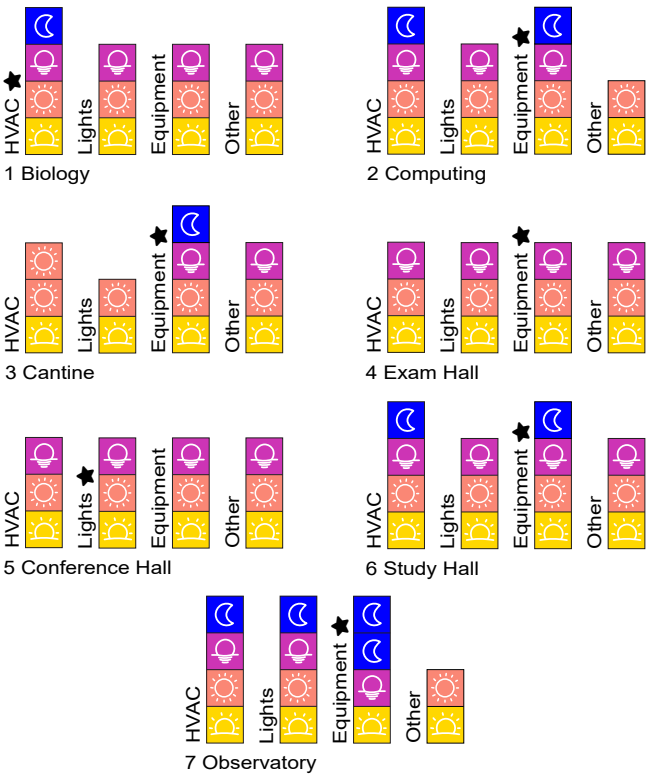


Figure 18: All goal profiles present in the simulation of the Energy Trails game. Main goal compositions are indicated with ★.

but the negotiation component checks which action was taken and (if applicable) calls on the next agent to take an action.

Finally, in the negotiation component of the simulation, the turn order is decided. In this case, the order in which agents place an offer is always the same. For example, if we consider a game where a ToM_0 agent is playing with a ToM_1 agent, the ToM_0 agent always starts. If we consider a game where a ToM_1 agent is playing with a ToM_0 agent, the ToM_1 agent always starts.

The Game This component of the simulation handles the game environment in which the negotiation takes place. The function of this component is to keep track of game-related information, initialise the game, and facilitate parts of the negotiation that are specifically Energy-Trails related.

The game-related information stored here is: the number of chip colours, the number of chips per player, the number of boards, the board size, the possible goal profiles, the goal profiles that the players have, the chip sets that the players start with, the total chips in-game, the full utility function, the boards themselves, all information related to the boards, and the offer history of each player.

This component also handles setting-up the game. This means that the boards are generated, the chips are generated and distributed over the players, each player is assigned a goal profile, the total number of possible offers is calculated, the possible paths on each board are determined and the full utility function is calculated.

The board and chip generation are done semi-randomly. When these are generated, the chip distribution is considered such that at least one of each chip colour is always present in the game and on the boards. The goal profiles are also assigned semi-randomly; the players cannot have the same goal profile.

During the game set-up, the number of possible offers is calculated using the offer codes. Afterwards, for each goal profile, the relevant paths on all boards are found; paths which count as either reaching the board's respective goal composition or taking a step towards the board's respective goal composition.

If there is no path that reaches the board's respective goal composition, the process is repeated with a "bonus action". This means the path can contain one tile that doesn't directly count for reaching the board's respective goal composition, as long as there are enough chips available in the game as a whole to facilitate this path. If, again, no path is found that reaches the board's respective goal composition, this process is repeated with an additional bonus action until a path is found that fully reaches the board's respective goal composition. This way, only the most efficient paths are used to find the score of each offer.

To find the best distribution of the chips an agent would receive if the offer is accepted over the boards in the game, the paths calculated in the previous step are used. Initially this is done using only the chips needed to fulfil the complete goal composition of each board, including potential bonus actions. For the offers possible with these chips, every single distribution of the offer over the possible paths on the board is considered and the score (excluding shared points) is calculated. In this stage, leftover chips are "discarded". This means that, initially, only the distributions of offers are calculated in which there are no leftover chips.

Then for every possible offer in the game, the highest-scoring "optimal" offer is found that can be made with the offer at hand, excess chips are labelled as leftover chips, and the score for this offer is set accordingly. As a result, a list of scores is made for every goal profile. This list keeps track of

the highest possible score per offer in the game, including how this offer would need to be distributed over the boards in the game. At the same time, for each offer, the shared score is calculated and stored separately. This shared score can be accessed by ToM_k agents with $k \geq 1$ when deciding which offer to place.

Overall, the utility function allows agents to quickly access the best score they can get for each offer that can be placed with the chips that are in-game, for every goal profile. This is used to decide which offers they want to place, but it also allows agents to have accurate models of how their trading partner might rank the offers depending on the goal profile they have.

The utility function uses the scoring function, which is also stored in this component of the simulation. This function calculates the score of an offer depending on the parameters of the game. If a cooperative version of the game is played, which includes point sharing, the shared points are also calculated. The scoring function also handles the prioritisation of certain boards according to the goal profiles of agents.

The game component also facilitates converting offer codes into offer lists and vice versa, finding the difference between two chip lists, and converting an offer from one player's perspective into another player's perspective.

The Agents In Energy Trails, ToM_0 agents function similarly to how they do in Coloured Trails. They mainly use the utility function to decide which offer is the best for them to make at that moment in time. Since they are not aware that their trading partners have goal profiles, or even that they are also trying to maximise their score, the way ToM_0 agents work is not affected much by the change from locations to goal compositions nor by the change from one to multiple boards. Additionally, since the utility function does not include the shared score they would get when their trading partner reaches a goal composition, they also do not use the shared scores in the cooperative version of Energy Trails. The final score that ToM_0 agents receive at the end of the game does include the shared scores, but they cannot use this information to update their beliefs since the game has already ended.

The only change that was made to ToM_0, which was also made to ToM_1 and ToM_2 agents, is that keep track of their offer history and use this information when deciding which offer to place next. When they calculate the value of each offer, the agents now assign a value of 0 to offers they have placed before. This was introduced to prevent negotiation cycles. In negotiation cycles, agents would keep placing the same offers even if they did not get accepted by their trading partner. In the past, this lead to the negotiations ending in time-outs more frequently than is desirable. This issue was resolved by assigning a value of 0 to previously placed offers.

To manage the increased time complexity in ToM_1 and ToM_2 agents, an optional shortcut was implemented. To choose which offer to place, ToM agents simulate the offer selection procedures of their trading partner for each offer that they consider placing. For this, they go through all the entire recursive procedure: the beliefs of their internal simulated trading partner are updated according to the hypothetical offer, the simulated trading partner goes through the entire process of accepting, rejecting, or placing a counteroffer based on this hypothetical offer, and the original agent assigns a value to the hypothetical offer based on the outcome of this process.

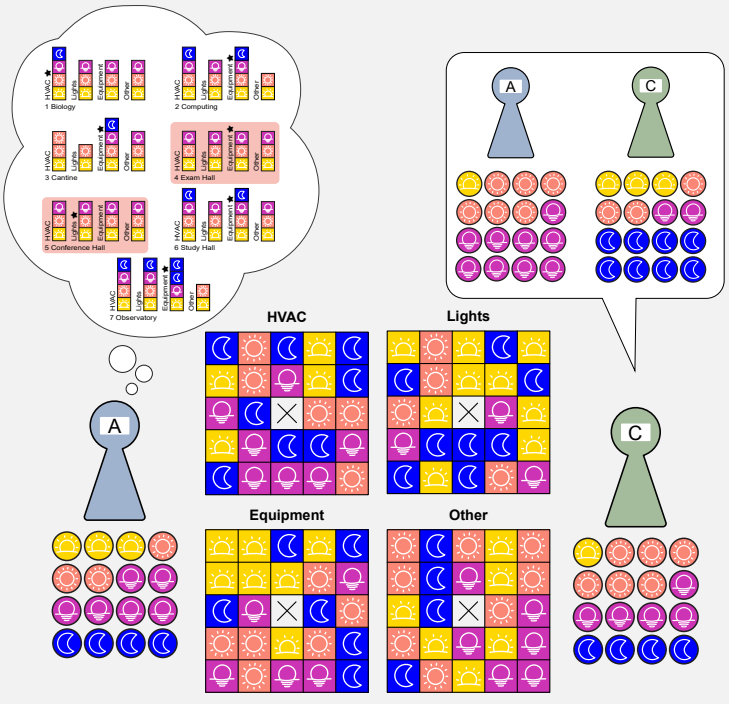
In order to make the process of offer selection less time-consuming, the first step of the simulated offer selection procedure can be skipped: the simulated trading partner does not update its beliefs based on the hypothetical offer. This means that instead of letting the simulated trading partner recalculate its offer values every time a hypothetical offer is placed, it uses its previously calculated offer values.

These values, to be clear, are not based on the actual beliefs of the ToM_2 agent's trading partner, but rather on what the ToM_2 agent believes those beliefs to be.

The benefit of this shortcut is best illustrated for ToM_2 agents. In a ToM_1 agent, complexity is not as much of a problem because its simulated trading partner is always a ToM_0 agent, which means that the simulated offer selection procedure is not very complex. For a ToM_2 agent, this is not the case. Since its simulated trading partner is most likely to be a ToM_1 agent, the simulated offer selection also contains simulated offer selections. Especially when Energy Trails is played on four boards, this process is very time-consuming.

Example 4 illustrates how a ToM_2 agent might act in the Energy Trails game. This example shows which beliefs are updated, just like Example 3, but not how the agent decides which beliefs to update. It also does not show the offer selection procedure of either agent, which is where the adjustments to the negotiation agents in our Energy Trails simulation as described above can be found. This shows that while we have made some adjustments to our negotiation agents, the overall way in which they function is still very similar to how they functioned in the simulation of Coloured Trails Weerd.

Example 4: First-order Theory of Mind in Energy Trails
Let us consider the situation illustrated below. Agent a is a ToM_1 agent, which means that it is trying to figure out the goal profile of agent c, l_c.
Upon receiving the offer as seen in the image below from agent c, it draws the following conclusions about the goal location of agent c: the locations 4 and 5 are unlikely to be the goal location of agent c because the offer by agent c assigns many night chips to agent c, and these goal profiles do not require any night chips. This means that, assuming that it still considered all goal profiles to be equally possible before this offer was placed, the number of goal profiles agent a considers possible for agent c has been reduced from 7 to 5.



3.3 Experimental Set-up

The performance of ToM_0 negotiation agents was compared to the performance of all possible pairings of negotiation agents containing at least one ToM_k agent with $0 < k \leq 2$.

The performance of ToM negotiation agents was tested under multiple different conditions, which can be found in Table 3. As explained in Section 3.1.4, we created two different versions of Energy Trails to reflect the real situation in the energy domain. Thus, we wanted to test the performance of ToM negotiation agents in both of these versions of Energy Trails.

The decision to test the performance of ToM negotiation agents in both a two-board version of Energy Trails and a four-board version of Energy Trails is related to the shortcuts we needed to make in the reasoning of ToM agents in Energy Trails due to the increased complexity. Since the complexity is lower in the two-board version of Energy Trails, we were able to use ToM_2 agents that reason without the reasoning shortcut. In the four-board version of Energy Trails, we had to use ToM_2 agents that reason with the shortcut.

By comparing the performance of ToM negotiation agents in these two conditions, we limit the effect of the shortcut on the overall results, while still being able to test the Energy Trails game with four boards since this is a better representation of the complexity of the negotiations about energy consumption.

Overall, the experiments step away from the building example explained in Chapter 1 and Section 3.1.1, but still attempt to model negotiations about energy consumption by buildings that aim to prevent overloading the energy grid.

For each of the negotiations, the boards were re-generated randomly and the agents were assigned new goal profiles from the list illustrated in Figure 18. During all negotiations between a certain agent pairing, the same agents were used. This means that ToM_0 agents were able to learn across negotiations.

As explained earlier, the negotiations each had a negotiation round limit of 25 due to time constraints. An example of what a negotiation looks like and which actions are counted as a negotiation round can be seen in Figure 19. In earlier work by De Weerd, Verbrugge, and Verheij [39] the negotiation round limit was more theoretical than practical because agents never negotiated until the negotiation limit. In Energy Trails, the negotiation round limit is practical rather than theoretical. The measures taken to prevent negotiation cycles were effective, but were not able to fully eliminate the occurrence of time-outs.

Number of Boards	Scoring	Shortcuts
2	Competitive	None
2	Cooperative	None
4	Competitive	ToM_2
4	Cooperative	ToM_2

Table 3: All conditions under which the performance of ToM_0 negotiation agents was compared to all possible pairings of negotiation agents containing ToM_1 and ToM_2 agents.

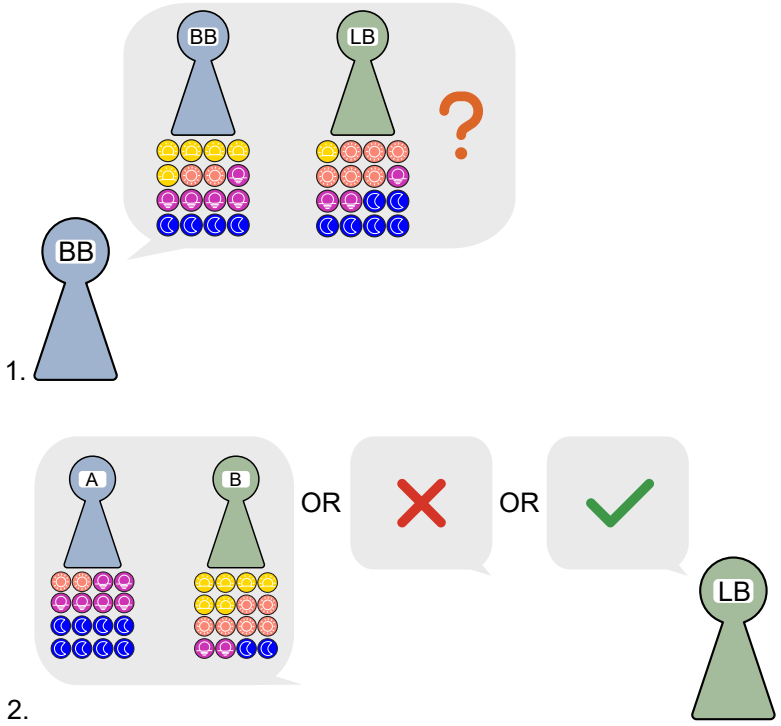


Figure 19: An example of an Energy Trails negotiation. If agent LB chooses the first option, which is to place a counteroffer, the game continues. Otherwise, the game ends.

Condition	ToM_0 vs. ToM_0	ToM_0 vs. ToM_1	ToM_0 vs. ToM_2	ToM_1 vs. ToM_1	ToM_1 vs. ToM_2	ToM_2 vs. ToM_2
2 boards, competitive	600	600	600	600	250	160
2 boards, cooperative	600	600	600	600	290	140
4 boards, competitive	125	125	125	25	50	125
4 boards, cooperative	125	125	125	20	40	125

Table 4: Table showing the number of negotiations in each experiment condition and for each agent-pairing. In each condition, the number of negotiations deviates for two agent pairings. These deviations are marked in red.

Since the negotiation speed, the time it took for a single negotiation to be finished, was not equal across conditions and agent-pairings, we were not able to run the simulation for a consistent number of negotiations across conditions. Instead, we opted to let the experiments run for as long as possible, and then cap the number of negotiations per condition such that the number of negotiations in the results of at most two agent-pairings deviates from the number of negotiations in the results of the other agent-pairings. The final number of experiments for each condition and agent-pairing can be found in Table 4.

4 Results

In this chapter, we look at the results of the experiments described in Section 3.3 by considering different performance metrics of theory of mind agents in a negotiation game.

In order to answer the research question, we will be comparing the performance of the ToM_0 vs. ToM_0 agent pairing with all other agent pairings with at least one ToM_k agent with $0 < k \leq 2$. This way, we can find if it is beneficial to use theory of mind in Energy Trails.

As described in Section 3.3, the experiments contained different numbers of negotiations in each condition. The specific numbers can be found in Table 4.

In all graphs, the agent pairings are indicated by their orders of theory of mind. For example, this means that the label "0 vs. 1" indicates a ToM_0 vs. ToM_1 agent pairing.

4.1 Acceptance Rate

The first metric we used to test the performance of theory of mind in negotiations was acceptance rate. The acceptance rate is the percentage of negotiations that ended because one of the agents accepted an offer from their trading partner.

Based on the shape of the bar graphs alone, which can be seen in Figure 20, we see that theory of mind has different effects in all the conditions. The graph of which the shape is most consistent with previous work [34] is that of two boards with competitive scoring, which can be seen in Figure 20a. In this graph, we see that the performance increases as the mean order of theory of mind of the agents in the negotiation increases, with the exception of the ToM_2 vs. ToM_2 agent pairing, where the acceptance rate is slightly lower than that of the previous combination.

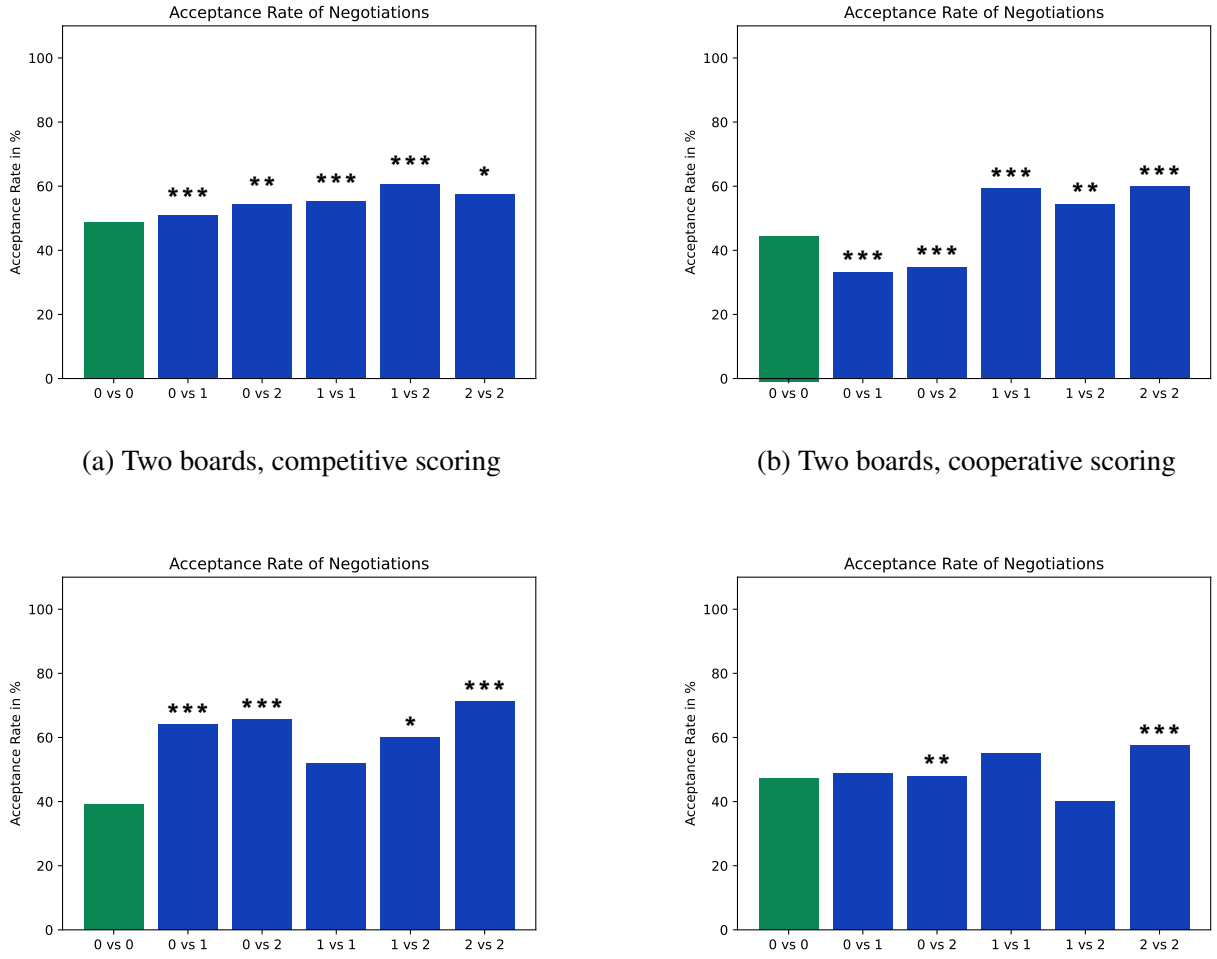
We find the largest difference between the ToM_0 vs. ToM_0 agent pairing and the other agent pairings when we play the Energy Trails game with four boards and a competitive scoring, as can be seen in Figure 20c.

Whether the difference between the acceptance rate of the ToM_0 vs. ToM_0 agent pairing was significantly different from the acceptance rate of each of the other agent pairings was tested with a χ^2 significance test and is also indicated in Figure 20.

Most notably, we find that the acceptance rates of the ToM_0 vs. ToM_1 and ToM_0 vs. ToM_2 agent pairings are significantly lower than that of the ToM_0 vs. ToM_0 pairing in the condition with two boards and cooperative scoring, as can be seen in Figure 20b.

In both conditions with competitive scoring, almost all pairings that contain at least one ToM_k agent with $k \geq 1$ have a significantly higher acceptance rate than the ToM_0 vs. ToM_0 pairing, as can be seen in Figure 20a and Figure 20c. The exception to this is the ToM_1 vs. ToM_1 agent pairing in the condition with four boards and competitive scoring.

Notably, whereas in both two board-conditions and in the four boards with competitive scoring condition almost all differences are statistically significant, in the four boards and cooperative scoring condition, only the acceptance rates of the ToM_0 vs. ToM_2 and ToM_2 vs. ToM_2 agent pairings are significantly different from the acceptance rate of the ToM_0 vs. ToM_0 pairing. This can be seen in Figure 20d.



(a) Two boards, competitive scoring

(b) Two boards, cooperative scoring

(c) Four boards, competitive scoring

(d) Four boards, cooperative scoring

Figure 20: The acceptance rate of the negotiations in all four conditions, where acceptance rate is the percentage of negotiations that were ended because either agent accepted an offer. Pairwise statistical significance between the results of the ToM_0 vs. ToM_0 agent pairing and all other agent pairings was tested with a χ^2 significance test and indicated with the following symbols: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Individual p-values can be found in Appendix A.

4.2 Score increase

The second metric we used to test the performance of theory of mind agents, is the mean score increase after negotiating, where score increase is the points that agents gained at the end of the game compared to the score they would have received for the initial chip distribution.

First, we will compare the results visually based on the bar plots as seen in Figure 21. Just like when comparing the acceptance rates, we find that the condition where the difference between the ToM_0 vs. ToM_0 agent pairing and the other agent combinations is the biggest, is in the condition with four boards and competitive scoring, as seen in Figure 21c.

We also find that in the condition with two boards and cooperative scoring, which can be seen in Figure 21b, the score increase seems to be higher when no ToM_0 agent is present in the negotiation.

The significance of these results was tested with pairwise comparisons between the ToM_0 vs. ToM_0 agent pairing and all other agent pairings using Dunn's test. The results of these tests are indicated in Figure 21.

We find that the mean score increase is significantly lower for the ToM_0 vs. ToM_1 agent pairing than the ToM_0 vs. ToM_0 agent pairing in the condition with two boards and cooperative scoring, as can be seen in Figure 21b. In this condition, we also find that the ToM_0 vs. ToM_2 agent pairing is the only pairing for which the mean score increase is not significantly different from the ToM_0 vs. ToM_0 agent pairing. This can be seen in Figure 21b.

We find that the two boards and competitive scoring condition is the only one in which all pairings containing at least one ToM_k agent with $k \geq 1$ have a significantly higher mean score increase than the ToM_0 vs. ToM_0 agent pairing, as can be seen in Figure 21a.

Lastly, we again find that the lowest number of significant differences is found in the four boards and cooperative scoring condition.

For these results, it is important to consider that the score increase is always zero when the negotiation did not end in an acceptance. This means that the mean score increase will heavily correlate with the acceptance rate. In order to separate this metric from the acceptance rate, we also consider the mean score increase after a successful negotiation, where a successful negotiation is one in which the score increased. This means that the negotiation ended in an acceptance.

When comparing the results visually, which can be seen in Figure 22, we find that removing the relation to the acceptance rates does seem to lead to different results. Bear in mind that only negotiations that ended in an acceptance were used to create these graphs, which means these graphs are based on a smaller number of data points than the other graphs in this chapter.

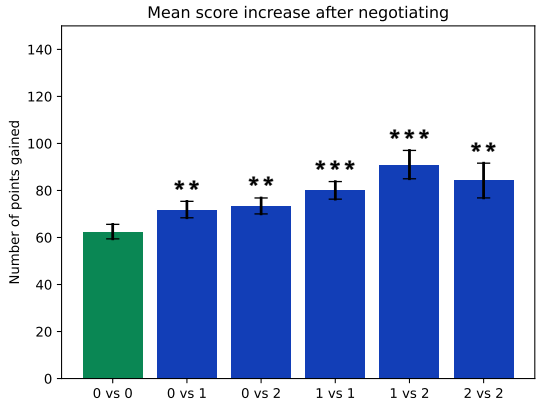
We find that the earlier effect of the presence of a ToM_0 agent seems to have diminished for the version of Energy Trails with two boards and cooperative scoring, as can be seen in Figure 22b. We also now find that the pairings which had a lower score increase than the ToM_0 vs. ToM_0 agent pairing, the ToM_0 vs. ToM_1 and ToM_0 vs. ToM_2 agent pairings, now have a higher score increase.

We still find that the biggest difference between the ToM_0 agent negotiation compared to the other agent pairings is overall the biggest in the four boards and competitive scoring version of Energy Trails, as can be seen in Figure 22c, but the difference is not as big as it was before. This can be seen in Figure 22c.

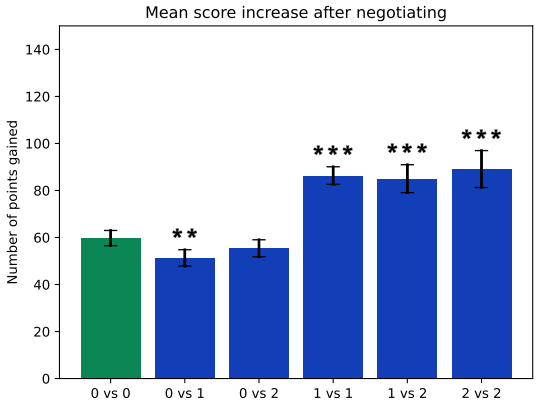
Like before, the significance of these results was tested with pairwise comparisons between the ToM_0 vs. ToM_0 agent pairing and all other agent pairings using Dunn's test. The results of these tests are indicated in Figure 22.

We find that compared to the mean score increase of all negotiations, where we found that in the two boards and competitive scoring conditions the ToM_0 vs. ToM_0 agent pairing had a significantly lower mean score increase than all other agent pairings, we now find that none of the other agent pairings have a score that is significantly different from the ToM_0 vs. ToM_0 agent pairing.

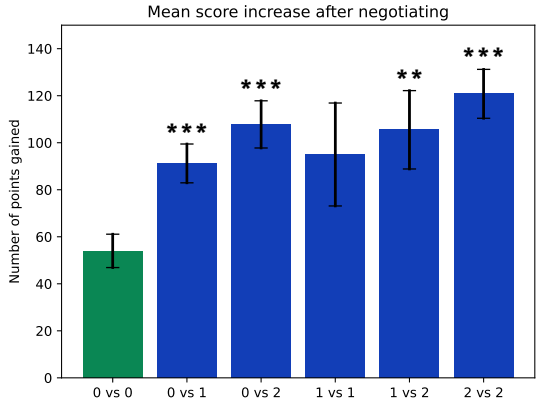
We also find a difference in the results of the two boards cooperative scoring condition for the ToM_0 vs. ToM_1 and ToM_0 vs. ToM_2 agent pairings, compared to the results for the mean score increase of all negotiations. We find that the mean score increase after successful negotiation of these agent pairings is significantly higher than that of the ToM_0 vs. ToM_0 agent pairing. This can be seen in Figure 22b.



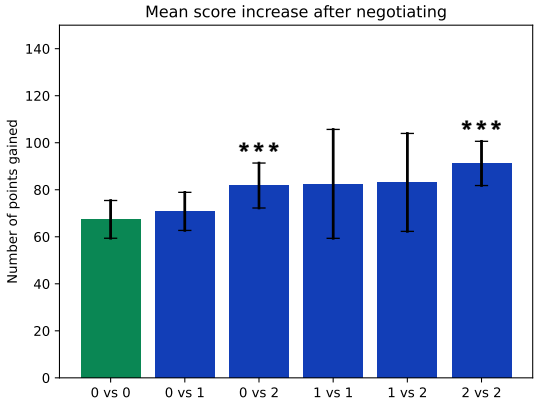
(a) Two boards, competitive scoring



(b) Two boards, cooperative scoring



(c) Four boards, competitive scoring



(d) Four boards, cooperative scoring

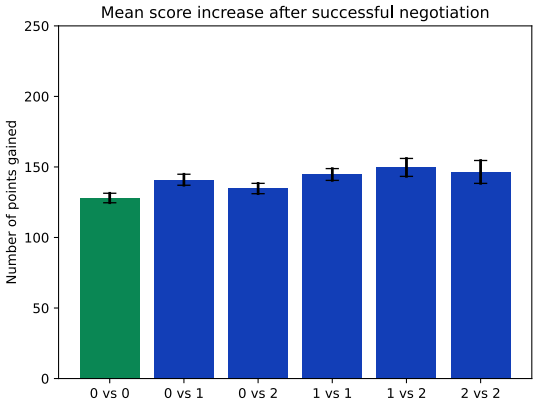
Figure 21: The mean score increase after negotiating, where score increase is the points that agents gained at the end of the game compared to the score they would have received for the initial chip distribution. Pairwise statistical significance test results using the Dunn's test between the ToM_0 vs. ToM_0 agent pairing and all other agent pairings are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Individual p-values can be found in Appendix A.

4.3 Negotiation Rounds

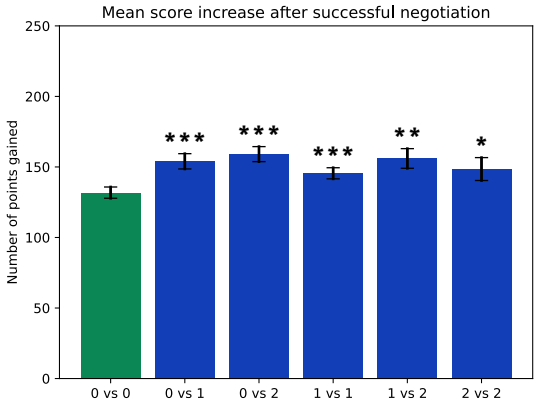
Another metric we used to test the performance of theory of mind in negotiation agents in the Energy Trails game is the number of negotiation rounds needed to end the game. Before we can consider this metric, however, we need to consider the time-out rate of each agent pairing.

The time-out rate, as seen in Figure 23, is the percentage of negotiations that ended because the number of negotiation rounds passed the limit rather than because either agent rejected or accepted an offer from their negotiation partner. The time-out rates are important to consider when looking at number of negotiation rounds as a metric, because they impact the mean heavily. The significance of these results was not tested.

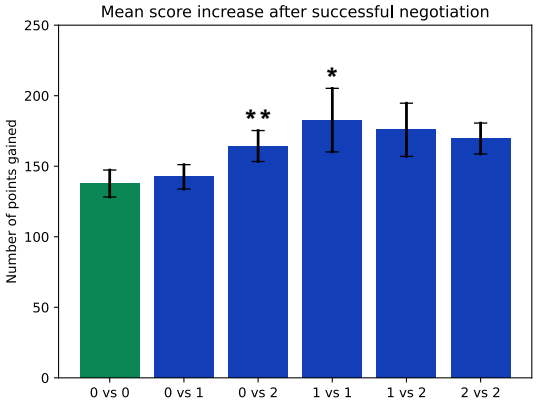
The plots as seen in Figure 23 all show that time-outs only occur for agent pairings in which no ToM_0 agent is present.



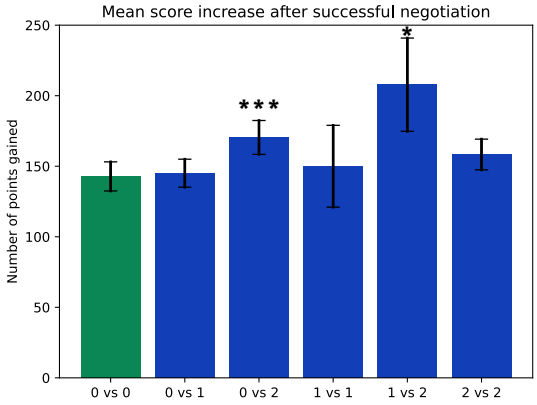
(a) Two boards, competitive scoring



(b) Two boards, cooperative scoring



(c) Four boards, competitive scoring



(d) Four boards, cooperative scoring

Figure 22: The mean score increase after a successful negotiation, which means the negotiation ended with either agent accepting the offer of their negotiation partner. Pairwise statistical significance test results using the Dunn's test between the ToM_0 vs. ToM_0 agent pairing and all other agent pairings are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Individual p-values can be found in Appendix A.

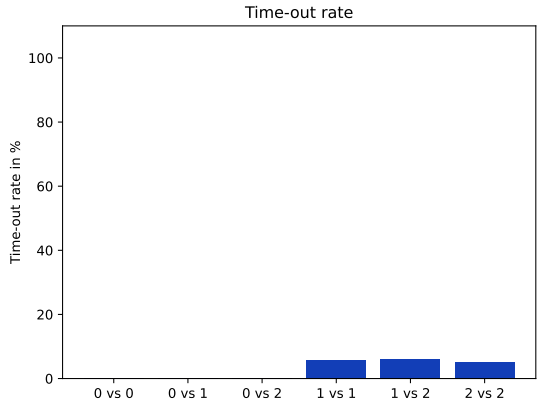
We see that the time-out rate increases with the increased mean order of theory of mind of agent pairings when Energy Trails is played with cooperative scoring and two boards, as seen in Figure 23b.

Conversely, we find that in the condition with four boards and competitive scoring, the time-out rate decreases with the increased mean order of theory of mind of agent pairings, as can be seen in Figure 23c.

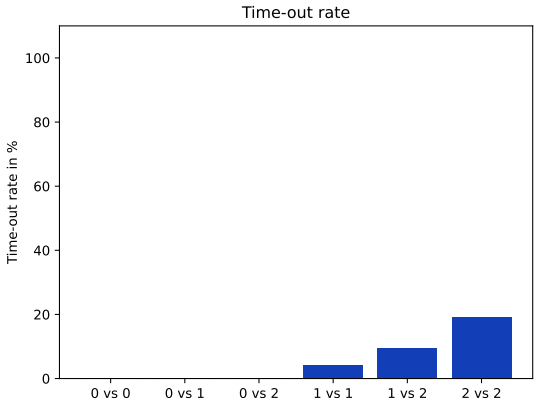
Overall, we also find that the time-out rates are higher in the versions of Energy Trails played with four boards, as can be seen in Figure 23c and Figure 23d.

In the condition with two boards and competitive scoring, we find that the overall time-out rate is the lowest and that it does not change much with the increased mean order of theory of mind, as can be seen in Figure 23a.

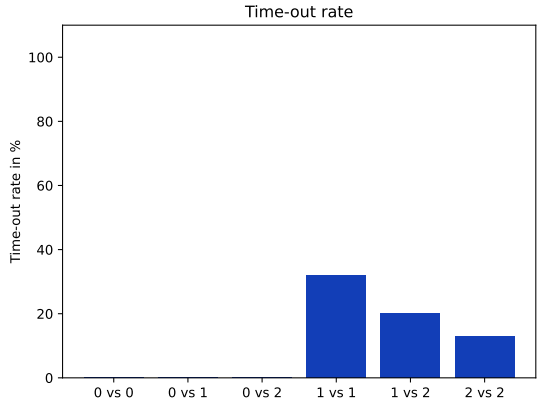
Now that we have considered the time-out rates of the different agent pairings in the different con-



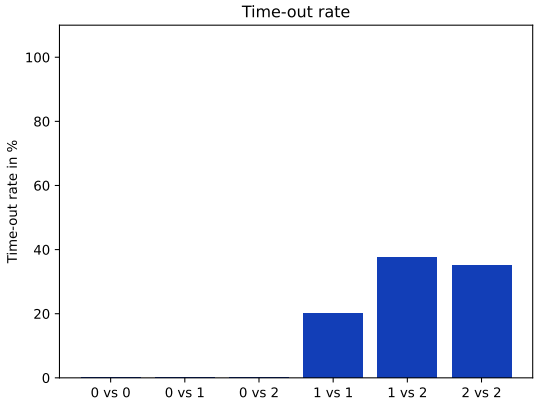
(a) Two boards, competitive scoring



(b) Two boards, cooperative scoring



(c) Four boards, competitive scoring



(d) Four boards, cooperative scoring

Figure 23: The percentage of negotiations that ended with a time-out rather than an acceptance of withdrawal.

ditions, it is time to consider the mean number of negotiation rounds. Here, we have excluded the time-outs because we considered those separately.

When considering the results as seen in Figure 24, we find the biggest differences when the game is played with cooperative scoring. In these cases, as can be seen in Figure 24b and Figure 24d, we find that the mean number of negotiation rounds is higher for agent pairings in which no ToM_0 agent is present.

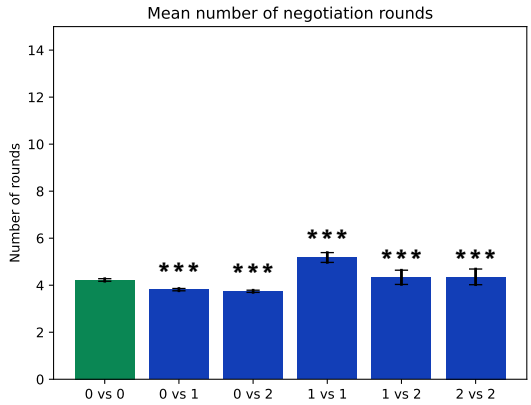
The highest overall mean number of negotiation rounds can be found in the condition with four boards and cooperative scoring for the ToM_1 vs. ToM_1 agent pairing, where the mean number of negotiation rounds is almost 12, as can be seen in Figure 24d.

In both conditions with four boards, we find that the mean number of negotiation rounds is higher for the agent pairings in which no ToM_0 agent is present. This can be seen in Figure 24c and Figure 24d.

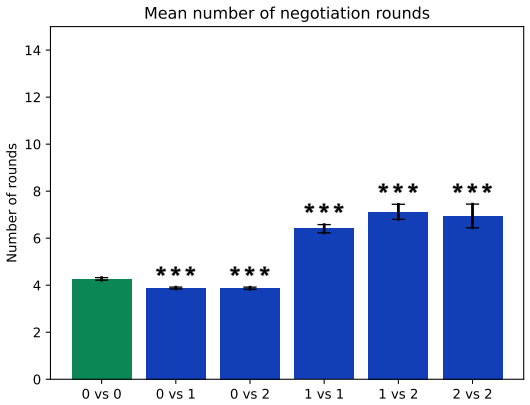
The significance of these results was tested with pairwise comparisons between the ToM_0 vs. ToM_0 agent pairing and all other agent pairings using Dunn's test. The results of these tests are indicated in Figure 24.

Using this test, we find that all results shown for the agent pairings in Figure 24 are significantly different from the results shown for the ToM_0 vs. ToM_0 agent pairing, except for the ToM_0 vs. ToM_1 agent pairing in the condition with four boards and cooperative scoring.

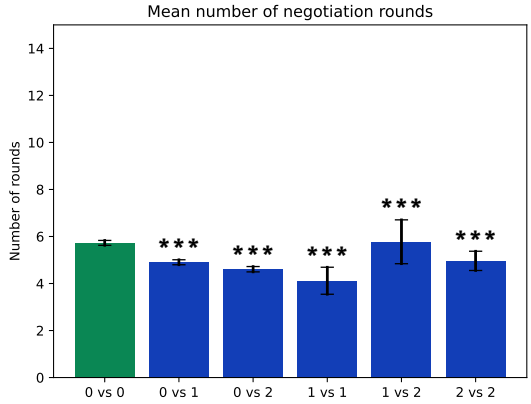
We also find that almost in all conditions, the number of negotiation rounds used is lower for the ToM_0 vs. ToM_1 and ToM_0 vs. ToM_2 agent pairings than for the ToM_0 vs. ToM_0 agent pairing. The only exception to this is the ToM_0 vs. ToM_1 agent pairing in the condition with four boards and cooperative scoring, which can be seen in Figure 22d.



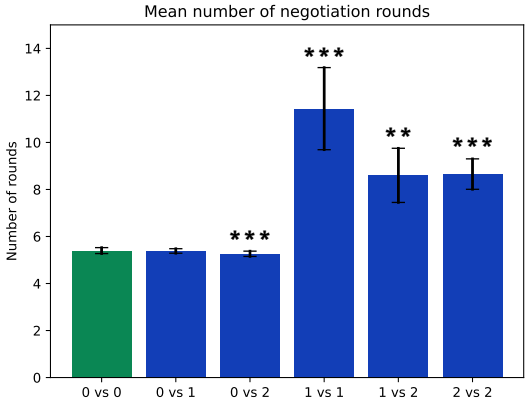
(a) Two boards, competitive scoring



(b) Two boards, cooperative scoring



(c) Four boards, competitive scoring

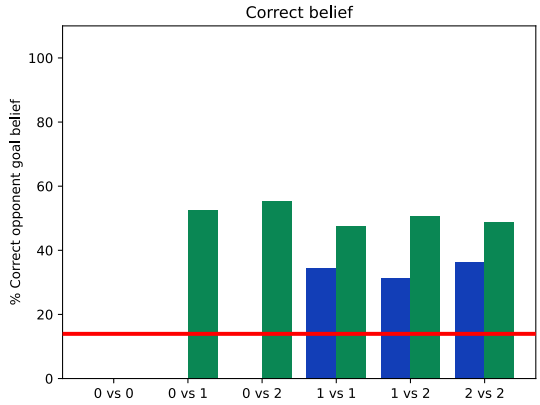


(d) Four boards, cooperative scoring

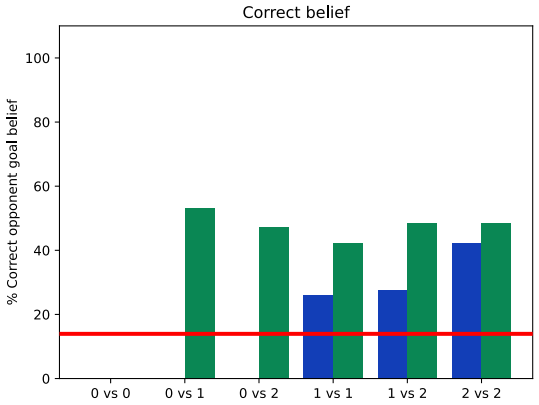
Figure 24: The mean number of negotiation rounds used by agents before the negotiation was ended with either an acceptance or a withdrawal. Pairwise statistical significance test results using the Dunn's test between the ToM_0 vs. ToM_0 agent pairing and all other agent pairings are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Individual p-values can be found in Appendix A.

4.4 Belief Correctness

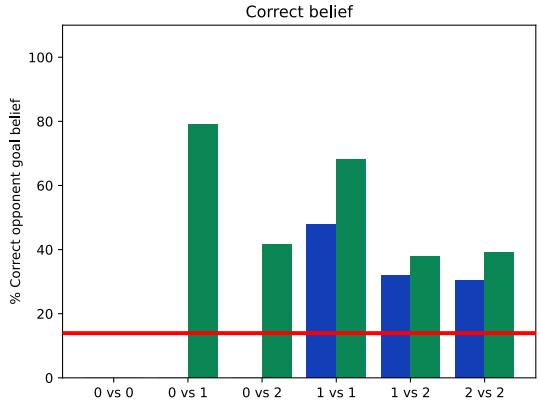
While the previous results focused on metrics for the performance of theory of mind in negotiations in Energy Trails, the belief correctness metric serves as a way to validate the performance of ToM_k agents with $k \geq 1$ to see if they perform better than a random guesser.



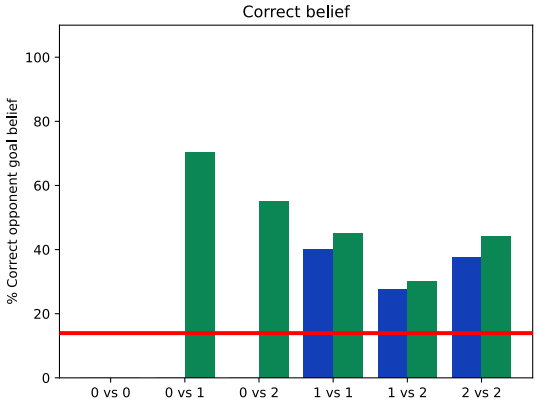
(a) Two boards, competitive scoring



(b) Two boards, cooperative scoring



(c) Four boards, competitive scoring



(d) Four boards, cooperative scoring

Figure 25: The percentage of games in which the goal profile that a ToM_k agent with $k \geq 1$ believed to most likely be the goal profile of their trading partner, was in fact the goal profile of their trading partner. A red line in each graph indicates the percentage of correctness we expect from a random guesser, which is based on the number of possible goal profiles. The bar on the left of the tick always indicates the starting agent, and the bar on the right of the tick always indicates the other agent.

To recap, the belief of an agent specifies, per potential goal profile of its trading partner, how likely it believes it to be that each goal profile is the goal profile of its trading partner. The beliefs per goal profiles add up to 1.

Thus, the metric of belief correctness is the percentage of games in which a ToM_k agent with $k \geq 1$ had a correct belief about the goal profile of their trading partner. Here, correct belief means that the goal profile with the highest belief corresponded to the actual goal profile of their trading partner.

Figure 25 shows the belief correctness of each ToM_k with $k \geq 2$ agent in each agent pairing. The bar on the left of the tick always indicates the starting agent, and the bar on the right of the tick always indicates the other agent.

Here, we find that the differences between agent pairings are not very different in both two board conditions, as can be seen in Figure 25a and Figure 25b.

In the four-board conditions, as can be seen in Figure 25c and Figure 25d, the differences between different orders of ToM agents seem to be bigger, but the differences are not consistent across these conditions.

We find that in all conditions, the starting agent has a lower belief correctness.

To compare these results against randomness, we assumed that a random guesser would be correct 1 in 7 times, since there are 7 possible goal profiles and goal profiles are assigned randomly. This percentage is indicated in each graph in Figure 25 with a red line.

We find that all ToM_k agents with $k \geq 2$ perform better than a random guesser. No significance tests were performed.

5 Discussion, Future Work & Conclusions

This section contains the discussion, explaining the strong and weak points of this thesis, the future work, outlining what research can be done based on the findings and process of this thesis, and finally, the conclusions of this thesis with regard to the main research question and all sub-questions as specified in Chapter 1.

5.1 Discussion

This section will outline the benefits of Energy Trails and our agent simulations, the improvements that can still be made, and the scientific relevance of this thesis.

5.1.1 Benefits

While the performance of using theory of mind was only tested through different performance metrics as seen in Chapter 4, there are more benefits to the system as a whole, including both Energy Trails and the simulation of the ToM negotiation agents.

One of those benefits is the way the system learns. Currently, most Artificial Intelligence-based techniques make heavy use of machine learning and deep learning. Both of these techniques rely on having large amounts of data available for training. Gathering the data required for such a procedure can be expensive, both in terms of time and in terms of money. In practice, this means that for many applications, the use of machine learning or deep learning simply is not feasible. Whereas most systems, like our ToM_0 agent, need to learn over the course of multiple training sessions, the higher-order ToM agents learn over the course of a single game. This allows this system to exhibit semi-intelligent behaviour even in the absence of historical data.

Another benefit to the game is the explainability of the system. The way the beliefs of agents are represented is fairly understandable to humans attempting to understand an agent's mental content. When this is combined with the utility function, it should be fairly understandable to a human why an agent decides to place a certain offer over another. This level of explainability should help with the trustworthiness of the system to its users.

5.1.2 Improvements

While a lot of care was put into ensuring the validity of the results presented in this thesis, choices always have to be made to limit the scope of the work. Thus, there are some things that could have been done better than how they were executed for this thesis. This section will outline those possible improvements.

Tuning First and foremost, there was an overall lack of tuning in the development process of the simulation of Energy Trails. Most parameter values were taken directly from the work of De Weerd, Verbrugge, and Verheij [39]. While this is a good foundation and should be considered as such, the goal of the current study was also to change and expand upon it, and with the changes that were made to turn Coloured Trails into Energy Trails, it is likely that the current parameters are not optimal.

An example of a component of Energy Trails that could use tuning is the scoring. Currently, the scoring tables as seen in Table 1 are largely based on the scoring in the work of De Weerd, Verbrugge, and Verheij [39] and on how we, in theory, want players to prioritise events in Energy Trails. This,

however, has not been subject to specific testing. Since the scale was increased a lot from Coloured Trails game, it can be assumed that the scoring would need to be adjusted accordingly.

To elaborate, in the Coloured Trails game, the highest scoring event is reaching a goal location, which is awarded with 500 points. Based on the total number of chips available in the game, a player would not be able to reach this score by having exclusively leftover chips. In the Energy Trails game, the highest scoring event is reaching a main goal composition, which is awarded with 1000 points. This means that a player needs 20 leftover chips to be awarded the same number of points as the highest scoring event in the game. Given that the total number of chips in the game is 32, this is not only possible, but also not very difficult. Especially considering that with four boards, at least some of the chips are expected to count towards the goal compositions of an agent, which awards more points than having leftover chips.

Negotiation Cycles Another aspect to be improved is how we handled the presence of negotiation cycles. In the absence of ToM_0 agents, it would often happen that agents got stuck in a negotiation cycle. What happens in such a cycle, is that one or both of the agents keep repeating the same offer or keep alternating between a small number of offers. This way, they would frequently reach the negotiation round limit. In the version of the simulation used for the experiments, this was prevented by keeping track of the agents' offer histories and assigning previously placed offers a value of 0.

While this was a successful way of preventing offer cycles, it did not address this issue at the source, nor did it fully eliminate the occurrence of time-outs. In the original simulation by Harmen de Weerd, this issue was prevented in negotiations with ToM_0 agents by only allowing the ToM_0 agents to decrease how strongly they believe a goal profile to be the goal profile of their trading partner as they observe new offers. A similar solution would have to be found to prevent negotiation cycles and time-outs from occurring in negotiations with higher-order agents.

Complexity Another improvement that can be made, of which the effect on the results is less clear-cut, is further optimising the time-complexity of the simulation. This can be done by increasing the efficiency of existing processes in the simulation, by doing a thorough check for redundant processes and by rewriting the simulation in a more efficient programming language, such as C. Additionally, the computations in the simulation as it is right now are very complete. All utilities for all goal profiles are computed beforehand, and the values of all offers are always considered.

This ensures that no situations are overlooked, but it also means that a lot of computations are done, of which it is almost certain that some are redundant. This can be avoided by integrating the use of certain heuristics, or even by using forms of machine learning to predict which offers may have high scores and should be considered. For the current research, using such an approach did not make much sense, as the goal of this research was to test the potential of theory of mind in the most optimal theoretical conditions.

Code Refinement Finally, in order to use ToM agents in negotiations about energy consumption in real buildings, the simulation needs to be made production ready. Currently, the code very clearly caters towards experiments and research. This needs to be changed to make it more clearly object-oriented and more generally applicable. This might also shed light on some small code-based oversights that can be at the root of issues such as offer cycles.

5.1.3 Scientific Relevance

While there are some improvements to be made to the system, we still believe that this thesis has scientific relevance. The results of the experiments done for this thesis show that theory of mind has potential for use in applied settings. Since previously, theory of mind has usually been used in theoretical research, this is a big step.

We found a way to create a bridge between the fundamental theory of mind research and the application of theory of mind in the energy domain. The game-theoretical model of negotiations in energy regulation seems to work with the theory of mind agents designed by Weerd, Verbrugge, and Verheij [39].

By doing this, we created a smart, explainable, and understandable system that can help prevent overload of the energy grid through automated negotiation. We, and some of the domain experts we spoke to, believe this is the next necessary step towards efficiently using our limited energy resources.

5.2 Future work

We believe that the use of Theory of Mind as described in this Master's Thesis has a lot of potential for both energy regulation and other domains. Thus, we have several ideas for future projects to add to or expand on the work described in this research.

One of the potential improvements is to find a way to adjust the Energy Trails game and our general simulation to work when there are more than two agents involved in the negotiation. One way to do this is by assigning the following roles to three agents: the allocator, the responder, and the competitor. In this hypothetical one-shot variation of Energy Trails, which would be based on the one-shot variation of Coloured Trails [34], the allocator and competitor simultaneously place an offer and the responder can decide whether to accept the offer of the allocator, accept the offer of the competitor, or reject all offers. Of course, other alternatives should be considered to find what works best with the Energy Trails game. The goal of this is to let a group of agents negotiate at the same time. This could be useful, for example, when you need all buildings on a campus to reach a consensus together rather than just two of them.

This ties into the next option for future work, which is to expand Energy Trails to represent negotiations on different scales. This means that the simulation needs to be generalised such that it would work for buildings amongst one another as well as areas in a building, or even different groups of buildings amongst one another. This would allow this system to be applied to a broader range of systems.

Expanding on the previous idea, future work could attempt to adjust Energy Trails, or Coloured Trails, further to fit domains other than distributing energy consumption to avoid overloading the energy grid. Previous work found that using theory of mind can be beneficial in many different settings [34], and this thesis was able to extend that finding to the energy domain. There are more domains that deal with incomplete information in cooperative, competitive, and mixed-motive settings, such as the domains of public safety and production lines, and thus it would be interesting to see if the findings of De Weerd [34] also extend to those other domains.

Another interesting perspective to consider for the current work is to test how our implementation of ToM agents compares to how humans use theory of mind. In that spirit, future work could focus on finding how this theoretical theory of mind we have created here compares to one that prioritises

being a realistic reflection of how theory of mind works in humans rather than on performance. An example is that in this study, the starting beliefs of ToM agents were not to believe that its trading partner had the same goal profile as itself, even though this is a strategy often used by humans.

Continuing on that point, another interesting future research topic could be to teach humans how to perform the negotiations as performed by ToM agents in this Thesis themselves. We already know that humans use theory of mind when playing a negotiation game against a ToM agent [6], so it would be interesting to see if this finding extends to humans playing the Energy Trails game with ToM agents. Additionally, humans might be able to learn from the mechanisms used by ToM agents to perform their own negotiations which, especially in a domain as complex as energy negotiations, might be advantageous.

Last but not least, an important step that can be taken in the future is to implement the system as described in this thesis into energy-regulating systems that are currently functioning in real buildings. In this process, humans would play an important role in terms of accurately describing the real situations in a simplified manner in order to be able as much of the Energy Trails framework as possible without running into complexity issues. During the time taken for this thesis, we spoke to domain experts and business in the energy domain who expressed a strong interest in using automated negotiations with theory of mind agents in their systems.

5.3 Conclusions

In order to answer our main research question **"How can theory of mind be used in computational models of negotiations in a Human-AI ecosystem, to be tested by way of a game?"** we answered the following sub-questions:

How can the Coloured Trails game be modified to correspond with real-life negotiations about energy consumption between two buildings? We changed the goals of agents from goal locations to goal profiles, consisting of multiple goal compositions, and allowed the agents to prioritise between sub-goals by introducing multiple boards to match the use-cases of the individual goal compositions. This led to a new game called Energy Trails.

How can the modified Coloured Trails be used to create a multi-agent system that simulates negotiations about energy consumption between two buildings? We created a multi-agent simulation based on the simulation of Coloured Trails by De Weerd [34]. Changes in this simulation reflected the changes made to Coloured Trails to create Energy Trails, and complexity problems were solved by introducing shortcuts.

Does using theory of mind provide an advantage in simulated negotiations about energy consumption between two buildings? We created a simulation of the Energy Trails game as described in Chapter 3 in which we let different orders of ToM agents negotiate with each other. We analysed and compared the performance of the different agent pairings across the conditions as described in Section 3.3 using different performance metrics.

For the first metric, the acceptance rate, we found that overall, the ToM agent pairings' performance was only worse when paired with a ToM_0 agent in the two boards conditions with cooperative scoring. In all other conditions and pairings, they were at least equal or better. We found that the beneficial effect was strongest using the competitive scoring.

For the second metric, the score gain, we considered the score gain both with and without negotiations that did not end in an acceptance. Again, we found that overall, the ToM agent pairings' performance was only worse when paired with a ToM_0 agent in the two boards condition with cooperative scoring. In all other conditions and pairings, they were at least equal or better. We found that the beneficial effect was strongest using the competitive scoring.

When we included the negotiations that did not end in an acceptance, we found that the benefit of the ToM pairings almost fully diminished in the conditions with competitive scoring. In all settings and across all agent pairings, we found the performance to be at least equal or better than the performance of the ToM_0 vs. ToM_0 agent pairing. We only found a consistent benefit for ToM agents in the condition with two boards and cooperative scoring.

For the third metric, the number of negotiation rounds used before the negotiation ended, we found that the pairings with one ToM_0 agent consistently performed at least equal to or better than the ToM_0 vs. ToM_0 pairing. In all conditions except for the one with four boards and competitive scoring, the pairings without ToM_0 agents consistently needed more negotiation rounds than the ToM_0 vs. ToM_0 agent pairing. Time-outs occurred in all conditions, and only for agent pairings without ToM_0 agent.

To validate these results, we tested whether the ToM_k agents with $k \geq 1$ performed better than a random guesser. We found that they did indeed perform better than a random guesser in all conditions.

From this, we can conclude that theory of mind can provide an advantage in simulated negotiations about energy consumption between two buildings. When considering the potential advantage for users of negotiation systems with theory of mind agents, we find that the conditions of the simulated negotiations impact the way that ToM agents provide an advantage, if any.

We found that the safest condition to use ToM agents in, is a condition with competitive scoring. Here, the ToM agents never provided a disadvantage. We did, however, also find that they mainly provide an advantage when it comes to whether they reached an acceptance in a negotiation, and how long it took to reach that acceptance. If we only consider the score increases, the ToM agents are not consistently better in this condition, even if they do not provide a disadvantage.

Across all conditions, we found that pairings without ToM_0 agents never provided a disadvantage, except when it came to the number of negotiation rounds. The advantage provided by these pairings was the biggest in conditions with cooperative scoring.

These findings lead us to the answer to our main research question, which also serves as the conclusion of this Master's Thesis.

How can theory of mind be used in computational models of negotiations in a Human-AI ecosystem, to be tested by way of a game? We find that theory of mind can be used in computational models of negotiations in a Human-AI ecosystem by letting them negotiate in the Energy Trails game. In this game, we find that ToM agents perform better than random guessers in terms of belief accuracy, and that they are also able to provide an advantage when it comes to several performance metrics.

Overall, this means that ToM agents can and should be applied to fit the needs of users in the energy domain, such that the environment allows them to perform best in the metric that is most important to the user.

References

[1] Z. Akata, D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerincx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen, and M. Welling, "A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence," *Computer*, vol. 53, no. 8, pp. 18–28, 2020.

[2] L. Barlassina and R. M. Gordon, "Folk psychology as mental simulation," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Summer 2, Metaphysics Research Lab, Stanford University, 2017. [Online]. Available: https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/.

[3] S. A. J. Birch and P. Bloom, "The curse of knowledge in reasoning about false beliefs," *Psychological Science*, vol. 18, no. 5, pp. 382–386, 2007.

[4] S. A. J. Birch, P. E. Brosseau-Liard, T. Haddock, and S. E. Ghrear, "A 'curse of knowledge' in the absence of knowledge? People misattribute fluency when judging how common knowledge is among their peers," *Cognition*, vol. 166, pp. 447–458, 2017.

[5] M. Blokpoel, M. van Kesteren, A. Stolk, P. Haselager, I. Toni, and I. Van Rooij, "Recipient design in human communication: Simple heuristics or perspective taking?" *Frontiers in Human Neuroscience*, vol. 6, p. 253, 2012.

[6] E. Broers, "Negotiating with incomplete information: The influence of theory of mind," M.S. thesis, 2014.

[7] S. Brok, "The influence of lying in a negotiation setting: Colored trails," M.S. thesis, 2023.

[8] R. W. Byrne and A. Whiten, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, English. Oxford: Clarendon Press; 1988.

[9] F. Charbonnier, T. Morstyn, and M. D. McCulloch, "Coordination of resources at the edge of the electricity grid: Systematic review and taxonomy," *Applied Energy*, vol. 318, p. 119 188, 2022.

[10] M. Deutsch and R. M. Krauss, "The effect of threat upon interpersonal bargaining." *The Journal of Abnormal and Social Psychology*, vol. 61, pp. 181–189, 1960.

[11] E. Erdogan, F. Dignum, R. Verbrugge, and P. Yolum, "Abstracting minds: Computational theory of mind for human-agent collaboration," in *HHAI2022: Augmenting Human Intellect*, IOS Press, 2022, pp. 199–211.

[12] S. Fatima, S. Kraus, and M. Wooldridge, *Principles of Automated Negotiation*. Cambridge: Cambridge University Press, 2014.

[13] S. Frey and R. L. Goldstone, "Cyclic game dynamics driven by iterated reasoning," *PLOS ONE*, vol. 8, no. 2, e56416, Feb. 2013.

[14] Y. Gal, B. J. Grosz, S. Kraus, A. Pfeffer, and S. M. Shieber, "Agent decision-making in open mixed networks," *Artificial Intelligence*, vol. 174, pp. 1460–1480, 18 2010.

[15] Y. Gal, B. J. Grosz, S. Kraus, A. Pfeffer, and S. Shieber, "Colored trails: A formalism for investigating decision-making in strategic environments," in *Proceedings of the 2005 IJCAI workshop on reasoning, representation, and learning in computer games*, 2005, pp. 25–30.

[16] A. S. Goodie, P. Doshi, and D. L. Young, "Levels of theory-of-mind reasoning in competitive games," *Journal of Behavioral Decision Making*, vol. 25, no. 1, pp. 95–108, Jan. 2012.

[17] A. Gopnik and H. M. Wellman, "Why the child's theory of mind really is a theory," *Mind & Language*, vol. 7, no. 1-2, pp. 145–171, 1992.

[18] E. Herrmann, J. Call, M. V. Hernández-Lloreda, B. Hare, and M. Tomasello, "Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis," *Science*, vol. 317, no. 5843, pp. 1360–1366, Feb. 2007.

[19] D. Hutto and I. Ravenscroft, "Folk psychology as a theory," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Fall 202, Metaphysics Research Lab, Stanford University, 2021. [Online]. Available: https://plato.stanford.edu/archives/fall2021/entries/folkpsych-theory/.

[20] S. Kass and K. Bryla, *Rock Paper Scissors Spock Lizard*, 1995. [Online]. Available: http://www.samkass.com/theories/RPSSL.html (visited on Feb. 22, 2023).

[21] H. Moll and M. Tomasello, "Cooperation and human cognition: the Vygotskian intelligence hypothesis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 639–648, Feb. 2007.

[22] J. von Neumann, "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, vol. 100, no. 1, pp. 295–320, 1928.

[23] S. E. Newman-Norlund, M. L. Noordzij, R. D. Newman-Norlund, I. A. C. Volman, J. P. de Ruiter, P. Hagoort, and I. Toni, "Recipient design in tacit communication," *Cognition*, vol. 111, no. 1, pp. 46–54, 2009.

[24] D. Pimentel, M. A. Whitecraft, Z. R. Scott, L. Zhao, P. Satkiewicz, T. J. Scott, J. Phillips, D. Szimak, G. Singh, D. O. Gonzalez, and T. L. Moe, "Will limited land, water, and energy control human population numbers in the future?" *Human Ecology*, vol. 38, pp. 599–611, 2010.

[25] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?" *Behavioral and Brain Sciences*, vol. 1, pp. 515–526, 4 1978.

[26] J. P. de Ruiter, M. L. Noordzij, S. Newman-Norlund, P. Hagoort, and I. Toni, "On the origins of intentions," in *Sensorimotor foundations of higher cognition*, Oxford University Press, 2007, pp. 593–610.

[27] S. Samuel, "A curse of knowledge or a curse of uncertainty? Bilingualism, embodiment, and egocentric bias," *Quarterly Journal of Experimental Psychology*, p. 17 470 218 221 132 539, Oct. 2022.

[28] T. C. Schelling, *The Strategy of Conflict: with a new Preface by the Author*. Harvard University Press, 1980.

[29] F. Trainer, "Can renewable energy sources sustain affluent society," *Energy Policy*, vol. 23, pp. 1009–1026, 1995.

[30] R. Verbrugge, "Logic and social cognition: The facts matter, and so do computational models," *Journal of Philosophical Logic*, vol. 38, no. 6, pp. 649–680, 2009.

[31] R. Verbrugge, B. Meijering, S. Wierda, H. Rijn, and N. Taatgen, "Stepwise training supports strategic second-order theory of mind in turn-taking games," *Judgment and Decision Making*, vol. 13, pp. 79–98, Jan. 2018.

[32] L. Vyogtsky, *Mind in Society*, M. Cole, V. Jolm-Steiner, S. Scribner, and E. Souberman, Eds. Harvard University Press, Feb. 1978.

[33] K. J. Warner and G. Jones, "The climate-independent need for renewable energy in the 21st century," *Energies*, vol. 10, pp. 1–13, 2017.

[34] H. de Weerd, "If you know what I mean: Agent-based models for understanding the function of higher-order theory of mind," Doctoral dissertation, University of Groningen, University of Groningen, 2015.

[35] H. de Weerd, E. Broers, and R. Verbrugge, "Savvy software agents can encourage the use of second-order theory of mind by negotiators," in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 2015, pp. 542–27.

[36] H. de Weerd, R. Verbrugge, and B. Verheij, "Higher-order theory of mind in the tacit communication game," *Biologically Inspired Cognitive Architectures*, vol. 11, pp. 10–21, 2015.

[37] H. de Weerd, R. Verbrugge, and B. Verheij, "Higher-order theory of mind is especially useful in unpredictable negotiations," *Autonomous Agents and Multi-Agent Systems*, vol. 36, p. 30, 2 2022.

[38] H. de Weerd, R. Verbrugge, and B. Verheij, "How much does it help to know what she knows you know? an agent-based simulation study," *Artificial Intelligence*, vol. 199-200, pp. 67–92, 2013.

[39] H. de Weerd, R. Verbrugge, and B. Verheij, "Negotiating with other minds: the role of recursive theory of mind in negotiation with incomplete information," *Autonomous Agents and Multi-Agent Systems*, vol. 31, no. 2, pp. 250–287, 2017.

[40] H. de Weerd, R. Verbrugge, and B. Verheij, "Theory of mind in the mod game: An agent-based model of strategic reasoning," in *Proceedings of the European Conference on Social Intelligence (ECSI-2014)*, A. Herzig and E. Lorini, Eds., vol. 1283, CEUR Workshop Proceedings, 2014, pp. 128–136.

[41] A. Whiten and R. W. Byrne, *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge: Cambridge University Press, 1997.

[42] M. Wooldridge, "Intelligent Agents: The Key Concepts," in *Multi-Agent Systems and Applications II*, V. Mařík, O. Štěpánková, H. Krautwurmová, and M. Luck, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 3–43.

Appendices

A Significance Tables

Condition \ Agent Pairing	ToM_0 **vs.** ToM_1	ToM_0 **vs.** ToM_2	ToM_1 **vs.** ToM_1	ToM_1 **vs.** ToM_2	ToM_2 **vs.** ToM_2
Two boards, competitive scoring	$< .001$	0.002	$< .001$	$< .001$	0.012
Two boards, cooperative scoring	$< .001$	$< .001$	$< .001$	0.002	$< .001$
Four boards, competitive scoring	$< .001$	$< .001$	0.35	0.010	$< .001$
Four boards, cooperative scoring	0.136	0.007	0.313	1.0	$< .001$

Table A.1: P-values resulting from the pairwise χ^2 significance test used in order to test whether the difference in acceptance rate between the ToM_0 vs. ToM_0 agent pairing and each other agent pairing was significant in each of the 4 experiment conditions.

Condition \ Agent Pairing	ToM_0 **vs.** ToM_1	ToM_0 **vs.** ToM_2	ToM_1 **vs.** ToM_1	ToM_1 **vs.** ToM_2	ToM_2 **vs.** ToM_2
Two boards, competitive scoring	0.002	0.006	$< .001$	$< .001$	0.005
Two boards, cooperative scoring	0.002	0.127	$< .001$	$< .001$	$< .001$
Four boards, competitive scoring	$< .001$	$< .001$	0.071	0.002	$< .001$
Four boards, cooperative scoring	0.074	$< .001$	0.226	0.586	$< .001$

Table A.2: P-values resulting from the pairwise statistical significance test results using the Dunn's test used in order to test whether the difference in mean score increase between the ToM_0 vs. ToM_0 agent pairing and each other agent pairing was significant in each of the 4 experiment conditions.

Condition \ Agent Pairing	ToM_0 **vs.** ToM_1	ToM_0 **vs.** ToM_2	ToM_1 **vs.** ToM_1	ToM_1 **vs.** ToM_2	ToM_2 **vs.** ToM_2
Two boards, competitive scoring	0.937	0.724	0.212	0.093	0.219
Two boards, cooperative scoring	$< .001$	$< .001$	$< .001$	0.002	0.04
Four boards, competitive scoring	0.887	0.008	0.047	0.116	0.132
Four boards, cooperative scoring	0.295	$< .001$	0.796	0.026	0.148

Table A.3: P-values resulting from the pairwise statistical significance test results using the Dunn's test used in order to test whether the difference in mean score increase after succesful negotiation between the ToM_0 vs. ToM_0 agent pairing and each other agent pairing was significant in each of the 4 experiment conditions.

Condition \ Agent Pairing	ToM_0 **vs.** ToM_1	ToM_0 **vs.** ToM_2	ToM_1 **vs.** ToM_1	ToM_1 **vs.** ToM_2	ToM_2 **vs.** ToM_2
Two boards, competitive scoring	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
Two boards, cooperative scoring	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
Four boards, competitive scoring	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
Four boards, cooperative scoring	0.187	$< .001$	$< .001$	0.001	$< .001$

Table A.4: P-values resulting from the pairwise statistical significance test results using the Dunn's test used in order to test whether the difference in mean negotiation rounds needed per negotiation between the ToM_0 vs. ToM_0 agent pairing and each other agent pairing was significant in each of the 4 experiment conditions.