How does the inclusion of signaling in the Mod Game affect the behavior of Theory of Mind agents?

Bachelor's Project Thesis

Atakan Tekparmak, s4017765, a.tekparmak@student.rug.nl,
Supervisor: Dr H.A. de Weerd

Abstract: This study is about the ability of agents to reason about the unrevealed mental content of other agents in competitive settings. This method of reasoning is called Theory of Mind and has been found to be effective in terms of social competence and competitive settings like negotiation. The study will focus on the Mod Game, a general sum n-player game where participants simultaneously choose a number from 1 to m each iteration to gain points, and will investigate the effect of Theory of Mind, coupled with signaling, on the emergent strategies and behavior in this game. The study will be conducted using virtual agents and by simulating two different versions of the Mod Game: one normal and the other one with signaling, where a part of agent population will be able to signal their chosen number to the other agents and the remaining agents will be able to receive and process these signals to make their decisions.

1 Introduction

Adaptive agents, for example humans, tend to explore new policies to try to outperform opponents in competitive settings. One possible approach is to reason about the unrevealed mental content of the other agents and act accordingly to their state of mind. This method of reasoning about the others' mental state is called *Theory of Mind* (Premack & Woodruff, 1978), which has been found to be important for social competence (Liddle & Nettle, 2006) and other settings like negotiation (de Weerd et al., 2017) and the *Mod Game* (Veltman et al., 2019), which will be further discussed in this paper.

Theory of Mind (ToM) is the ability to understand that other people have their own thoughts, beliefs, and intentions that may be different from one's own. It is a key component of social cognition, which is the ability to understand and interpret the actions and intentions of others. Theory of Mind is a crucial skill for successful social interactions, as it allows individuals to take the perspective of others, predict their behavior, and respond appropriately. Theory of Mind is a skill that develops over time, and is typically fully developed by the age of 5 (Call & Tomasello, 2008). However, it may not be limited to humans, and has been claimed in other species such as chimpanzees (Premack & Woodruff, 1978).

In competitive settings, Theory of Mind can be an important factor in predicting and understanding the behavior of one's opponents. This is also supported by the Machiavellian Intelligence Hypothesis, which states that the evolution of intelligence is driven by the need to outwit others (Whiten & Byrne, 1988), as social success (achieved mainly using methods such as deception and manipulation, which requires a higher level of Theory of Mind) leads to reproductive success (Gavrilets & Vose, 2006), which is one of the most driving factors of evolution. For example, individuals with a well-developed Theory of Mind may be able to anticipate their opponents' moves in a game of Rock-Paper-Scissors, giving them an advantage over their opponents who use Theory of Mind to a lesser extent(De Weerd et al., 2013). Additionally, individuals with a strong Theory of Mind may be better able to interpret their opponent's non-verbal cues and respond accordingly.

Overall, Theory of Mind is an important aspect of social cognition that can play a significant role

in competitive settings not only for humans but for artificial agents as well. Theory of Mind can also be applied recursively, in the sense that an agent can reason about the Theory of Mind of another agent, and so on. This is sometimes mentioned as *Higher Order Theory of Mind*, where each consecutive order adds another layer of reasoning on top of the previous one. In De Weerd et al. (2013) it is shown that in many different competitive settings, higher orders of Theory of Mind have a competitive advantage over lower levels, but it should be noted that the competitive advantage between levels decreases as the levels go up and tests were done for only up to the fourth level Theory of Mind.

A particular competitive setting that such behaviour could be observed is The Mod Game, which, as described in Frey & Goldstone (2013) and Veltman et al. (2019), is a general sum n-player game in which the participants simultaneously choose a number k in the range $\{0, \ldots, m\}$ where $m, n > 0$. In a general sum game, the total value of the payoffs for all players does not have to add up to a fixed number. After a player chooses number k, they gain one point for each other player that chose $k - 1$ as their number. For example if a player chose 8, for each opponent that chose 7 they gain one point. One notable exception is the case of 0, in which players gain one point for each player that chose m. The Mod Game is similar to other competitive games used in Theory of Mind settings, such as Rock, Paper and Scissors in the sense that actions have dominance over another (choosing 17 "dominates/beats" choosing 16 and Rock dominates/beats Scissors). However, the Mod Game is different in the sense that it is a general sum game, and the payoffs for each player do not have to add up to a fixed number.

We will simulate a version of the Mod Game with signaling. In this version of the game, there will be two types of agents: signalers and receivers. The signaling agents will use their beliefs and intentions (which are updated every iteration using their chosen $\{Signal, Action\}$ pair and the choices of other agents in that iteration) to choose a $\{Signal, Action\}$ pair. They will then signal their chosen Signal to the receiving agents. The receiving agents will receive the signals and use their beliefs and intentions (which are updated every iteration using their chosen action, the signals received, and the choices of other agents in that iteration) to choose an Action. Finally, both the signaling agents and the receiving agents will act, and their choices will be registered for that turn. The scores of the agents and their beliefs and intentions will be updated accordingly.

The signaling in this scenario is a perfect example of *Cheap Talk*, which is a form of communication that is cost-free and does not affect the payoffs of the agents directly. In Farrell & Rabin (1996), it is shown that even if there is a small incentive to lie in a setting where cheap talk is possible, cheap talk can convey information and alter the Nash equilibrium. In the study it is also shown that in many discussed situations when the signals are not "self-committing", they do not alter the Nash equilibrium. The study has arguments for both sides of the spectrum, but it also suggests that in settings like Prisoner's Dilemma (Kuhn, 2008) most authors and scientists emphasize on the fact that the prisoners cannot communicate, as with even semi-credible cheap talk, the prisoners can change the overall game, which would lead to a new Nash equilibrium. This can be exemplified by Prisoner 1 saying "If you snitch on me, I'll kill you" to Prisoner 2 and being believed. Even though that statement from Prisoner 1 is not self-committing, (Osborne & Rubinstein, 1994), Prisoner 2 would have to consider that in their decision making process.

This version of the Mod Game with signaling was chosen to be investigated in order to explore the effect of theory of mind on the emergent strategies/behaviour previously discussed in this setting (Veltman et al., 2019), as well as whether the addition of signaling leads to increased cooperation, increased deception particularly by the receiving agents, or any other notable differences in agent behaviour. Specifically, we will examine how the inclusion of signaling affects the choices and strategies used by the agents, and whether it leads to any significant changes in their behavior. This particular setting is chosen over other possible candidates like Prisoner's Dilemma (Kuhn, 2008), because it could possibly lead to a cooperative outcome where players alternate in getting a point for themselves and willingly conceding a point to others. By studying the strategies used by the signaling agents to choose their Signal, Action pairs and the strategies used by the receiving agents to choose their actions based on the previous actions and signals received, we hope to shed light on our research question and

better understand the role of theory of mind in this version of the Mod Game with signaling. Another research goal is to see if the addition of cost-free signaling leads to significant behavioral differences compared to the agent behaviour and scores on the regular simulation.

2 Methodology

The main experiment methodology of this paper will involve simulating the Mod Game with and without signaling, using artificial agents. Different initial populations of agents will be used in the simulation, with varying levels of Theory of Mind as described in Premack & Woodruff (1978) and Liddle & Nettle (2006). The behavior of the agents will be observed and analyzed through and across the simulations to investigate the effect of different order configurations and the addition of signaling on the behavior of Theory of Mind agents in the Mod Game, as described in Veltman et al. (2019) and Frey & Goldstone (2013).

To begin, a computer program will be created to simulate the Mod Game with and without signaling. The program will be written in the Python (Van Rossum & Drake Jr, 1995) programming language, following the implementation of the Mod Game as described in Veltman et al. (2019). The program will be designed to simulate the game with any given number of agents (initial population configurations), and will include functionality for agents to signal and receive signals in the signaling simulation, on top of the general ability of being able to update their beliefs and intentions at every iteration/round that is present in the non-signaling version.

The simulations, one regular (without signaling) and the other one with signaling, will be run multiple times, each with a different initial population of agents. The initial populations will be varied in terms of the levels of Theory of Mind of the agents, as described in De Weerd et al. (2013). Each initial configuration of agents will be simulated 10 times and their results will be aggregated to obtain more sound results. For this version of the Mod Game, we chose $m = 22$.

2.1 The Regular Simulation

In the regular simulation, which is a basic n-player version of the Mod Game with Theory of Mind agents, three different processes are performed in each iteration/round of the simulation. These processes are:

1. The agents **decide** based on their beliefs,
2. The agent scores are updated, and
3. The agents beliefs are **update**d.

The **decide** and **update** processes are performed differently for each level of Theory of Mind. However, there are similarities between the processes, as per the requirement for a higher level Theory of Mind: accurately modeling the lower levels of Theory of Mind. How each order of Theory of Mind agent performs these processes is described below.

2.1.1 Zero Order Theory of Mind Agent

The Zero Order Theory of Mind agent is the simplest agent in the simulation. It does not have any Theory of Mind, and tries to model the other agents as simple as possible through a 1-dimensional belief vector that represents the agent's belief in most common action chosen by the other agents. It simply uses this vector to choose an action and update its beliefs. The **decide** function can be described as follows:

Let $\mathbf{b} = [b_0, b_1, \ldots, b_{22}]$ be the belief vector, where b_i represents the belief in any given opponent choosing action i. The function iterates through the beliefs and finds the index i^* corresponding to the highest belief:

$$i^* = \arg\max_i b_i$$

To incorporate exploration, a random action is chosen with probability ϵ using the epsilon-greedy strategy. The function employs the check_epsilon function to determine if a random choice should be made. The zero order decision d_0 is then given by:

$$d_0 = \begin{cases} \text{random\_choice}(), & \text{if check\_epsilon}() \\ i^* + 1 \mod 23, & \text{otherwise} \end{cases}$$

Here, random_choice() generates a random action, and $i^* + 1 \mod 23$ ensures that the action is within the range of 0 to 22.

The **update** function of the Zero order Theory of Mind Agent updates the belief vector, using the actions of other agents $\mathbf{a} = [a_0, a_1, \ldots, a_{n-1}]$ where n is the number of agents, as follows:

Let $\mathbf{b} = [b_0, b_1, \ldots, b_{22}]$ be the belief vector, where b_i represents the belief in action i. First, the function computes the accumulated sum of beliefs, denoted as A, using the equation:

$$A = \sum_{i=0}^{22} b_i.$$

Next, it calculates the learning rate α as:

$$\alpha = \frac{1 - \text{LEARNING\_SPEED}}{A}.$$

The learning rate α represents the proportion of belief update for each action. A higher accumulated sum of beliefs leads to a smaller learning rate, promoting stability and slower belief adjustments. Finally, the function updates each belief b_i by multiplying it with the learning rate α, except for the belief corresponding to the chosen action. The updated belief vector is given by:

$$b_i = b_i \times \alpha, \quad i = 0, 1, \ldots, 22,$$
$$b_{a_j} = b_{a_j} + \text{LEARNING\_SPEED},$$

where a_j represents an action from the list $\mathbf{a}$ that represents the list of actions of the other agents in that round. This process is repeated for each action in the actions list. One notable feature of this update equation is that it is performed sequentially according to the initialization order of the agents, so the order of the agents can affect the beliefs of the agents. The update equation ensures that the belief corresponding to the chosen action is increased by the predefined learning speed LEARNING_SPEED, while the other beliefs are scaled down proportionally.

2.1.2 First Order Theory of Mind Agent

The First Order Theory of Mind agent is a more complex agent than the Zero Order Theory of Mind agent. It models the other agents in the simulation as Zero Order Theory of Mind agents, uses that model in deciding and updating. They, like the zero order agents, have a fixed order Theory of Mind

The **decide** function of the First Order Theory of Mind Agent models the decision-making process of a lower-order agent (Zero order Theory of Mind Agent) and acts greedily based on that model. First, the function invokes the **decide** function of the lower-order agent, which determines the action the lower-order agent would choose based on its beliefs and decision-making strategy. Next, the function makes its decision $\mathbf{d_1}$ by applying a greedy strategy based on the lower-order decision $\mathbf{d_0}$:

$$d_1 = (d_0 + 1) \mod 23.$$

Because the First Order agent using a Zero Order agent to decide, the **update** procedure is just updating that Zero Order agent, as described in the previous subsection. It should be noted that unlike Veltman et al. (2019) first order agents are fixed to the first order of thinking, and cannot reason in a zero order way.

2.1.3 Second Level Theory of Mind Agent

The Second Order Theory of Mind agent is very similar to a First Order Theory of Mind agent, with the main difference being that it has another dimension of reasoning, namely its beliefs on which order of Theory of Mind is more present in the population, Zero Order or First Order. It then uses this belief to decide which lower order agent to model and act upon.

The **decide** function of the Second Order Theory of Mind Agent implements a greedy decision-making process based on models of the other agents that it has. First, the function selects an order that it believes to be more present in the population. Let $\mathbf{b_o} = [b_{o_1}, b_{o_2}]$ be the order belief vector. Then, the intermediary order-decision o^* is:

$$o^* = \begin{cases} 0, & \text{if } b_{o_1} > b_{o_2} \\ 1, & \text{otherwise} \end{cases}$$

In the order decision procedure there is epsilon-based stochasticity as well, the final order decision o is selected with:

$$o = \begin{cases} \text{random\_choice}(), & \text{if check\_epsilon}() \\ o^*, & \text{otherwise} \end{cases}$$

where *random_choice()* randomly chooses between 0 and 1. The agent then invokes the **decide**

function of the corresponding lower-order agent to get the decision d_l of the lower-order agent. Then, the second order decision $\mathbf{d_2}$ is calculated using the equation:

$$d_2 = (d_l + 1) \mod 23.$$

The **update** function of the Second Order Theory of Mind Agent updates the order beliefs based on the observed actions. First, the function retrieves the decisions made by the first order agents using their **decide** functions. The higher order decisions $\mathbf{d_{h_0}}$ and $\mathbf{d_{h_1}}$ are calculated using the equations:

$$d_{h_0} = d_1, \quad d_{h_1} = (d_1 + 1) \mod 23.$$

If the chosen action matches either the zero order higher decision $\mathbf{d_{h_0}}$ or the first order higher decision $\mathbf{d_{h_1}}$, the order beliefs $\mathbf{b_o}$ are updated as follows:

Let a be the observed action. The sum of the order beliefs is calculated as $s = b_{o_0} + b_{o_1}$, and the value of s is used to calculate an accumulation factor f:

$$f = \frac{1.0 - \text{LEARNING\_SPEED}}{s}.$$

The belief values in the order beliefs vector are scaled down by multiplying each belief b_{o_i} by the accumulation factor f. If a matches the zero order higher decision $\mathbf{d_{h_0}}$, the belief in the majority of zero order agents is updated by:

$$b_{o_0} = b_{o_0} + \text{LEARNING\_SPEED}.$$

Otherwise, if a matches the first order higher decision $\mathbf{d_{h_1}}$, the belief in the majority of first order agents is updated by:

$$b_{o_1} = b_{o_1} + \text{LEARNING\_SPEED}.$$

In all other cases, the order beliefs are not updated. Finally, the belief vector $\mathbf{b}$ is updated as described in Section 2.1.1.

2.1.4 Simulation

The non-signaling simulation will have a total of 7 different initial population configurations to be simulated, which are listed in Table 2.1. These population configurations are chosen specifically because for each ToM order it investigates the effects of:

- The Equal Population: $\{33\%, 33\%, 33\%\}$,
- The Surplus of an Order: $\{50\%, 25\%, 25\%\}$ and
- The Deficiency in an Order: $\{20\%, 40\%, 40\%\}$

0 Order	1st Order	2nd Order
33	33	33
50	25	25
20	40	40
25	50	25
40	20	40
25	25	50
40	40	20

Table 2.1: The initial population configurations listed by how much each order of ToM is present in the population percentage-wise

2.2 The Signaling Simulation

In the signaling simulation, the agents are divided into two groups: the signaling agents and the receiving agents. The signaling agents can signal any action (a number in the range $[0, 22]$) to the receiving agents, before deciding on an action to perform. The receiving agents can process transmitted signals from the signaling agents before deciding. There are five different processes performed in each round. In order, they are:

1. The signaling agents decide on a signal and transmit it
2. The receiving agents receive the signal and process it
3. All agents decide on an action
4. The agent scores are updated
5. The agent beliefs are updated

Even though not mentioned in the itemized list above, the decision and update processes for signaling and receiving agents are different. They do resemble in certain aspects as similar learning algorithms were used for consistency, but the differences are significant enough to warrant a separate explanation, which will be done in the following subsections.

2.2.1 Signaling Agents

All signaling agents employ a 2-dimensional beliefs array of the shape (23,23), in which the first dimension represents and holds information for signals, and the second dimension represents and holds information for actions given the signal. In the decision process for all signaling agents, the agent uses the sub-array of shape (23,1), indexed by the signal, to decide on an action. How each order of ToM deals with the signal, decision and update processes are explained in the following sections.

2.2.1.1 Zero Order Signaling Agent

The zero order agent employs a simple greedy strategy for signaling and deciding. In both of them, the signal tendency matrix **t**, which is randomly initialised for each individual agent, is used. It should be noted that while the belief vector **b** used by the regular agents as described in Section 2.1.1 represents the agent's belief in what the other agents will do in the next round, the signal tendency matrix **t** represents what the agent should do itself to maximise the point gain in the next round:

$$\mathbf{t} = [[t_{0_0}, \ldots, t_{0_{22}}], \ldots, [t_{22_0}, \ldots, t_{22_{22}}]]$$

where $\mathbf{t_{i_j}}$ is the agent's tendency to signal action **i** and then perform action **j**. The **signal** ϕ and the **decision** $\mathbf{d_0}$, which is chosen by the agent using the signal ϕ is given by:

$$\begin{aligned} \phi &= \arg\max_s \sum_{i=0}^{22} t_{s_i}, \quad i = 0, 1, \ldots, 22. \\ d_0 &= \arg\max_d t_{\phi d}, \quad d = 0, 1, \ldots, 22. \end{aligned} \tag{2.1}$$

After this, in the update part, when the agent observes the actions $\mathbf{a_j}$ that have been performed in the previous round, it updates its signal tendencies so that:

$$t_{i_k} \leftarrow t_{i_k} + LEARNING\_SPEED \tag{2.2}$$

where $k = (a_j + 1) \mod 23$ is the action that would have gained a point from the observed action $\mathbf{a_j}$ in the previous round. This update is performed sequentially for each observed action $\mathbf{a_j}$, and for each signal $0 \leq i \leq 22$. Note, however, that the signal tendencies are not normalized. That is, the order in which updates are performed does not influence the outcome.

2.2.1.2 First Order Signaling Agent

The first order agent employs a similar but slightly different strategy for signaling and deciding. In both of them, the same signal tendency matrix **t** is used. The agent also has a zero order agent model $\mathbf{m^0}$, whose **signal** ϕ (calculations given in Equation 2.1) is used in the signaling process. The **signal** $\mathbf{s^1}$ and the **decision** $\mathbf{d_1}$ chosen by the agent is given by:

$$s^1 = \begin{cases} \text{random\_choice}(), & \text{if check\_epsilon}() \\ s^0, & \text{otherwise} \end{cases}$$
$$d_1 = \arg\max_d t_{s^1 d}, \quad d = 0, 1, \ldots, 22.$$

In the **update** process, first the zero order agent model $\mathbf{m^0}$ is updated for all observed actions $\mathbf{a_j}$, but with $k = (a_j + 2)$ instead of $(a_j + 1)$ on the right hand side of the equation. This change in k is because the first order agent expects other agents to update their beliefs in response to the observed actions. Then, for all indexes **i** and **j** in the signal tendency matrix where $0 \leq i, j \leq 22$, the following algorithm is employed for update, where $\mathbf{l_s}$ denotes the LEARNING_SPEED:

$$t_{i_j} = \begin{cases} t_{i_j} + l_s \cdot (1 - t_{i_j}), & \text{if } i = s^1 \,\&\, j = d_1 \\ t_{i_j} \cdot (1 - l_s), & \text{otherwise} \end{cases}$$

It's important to note that his has the effect of increasing the tendency $t_{s^1 d_1}$ of choosing the same signal s^1 and making the same decision d^1 again in the future.

2.2.1.2 Second Order Signaling Agent

The second order signaling agent is a lot like a regular second order agent, with the lower order signaling agent models $\mathbf{m^0}$ and $\mathbf{m^1}$, and the order beliefs $\mathbf{b_o}$. The **decide** and **update** processes are the same as the Non-signaling Second Order Agent, with the only difference being the extra **signal** $\mathbf{s^2}$ chosen by the agent, given by the same algorithm

for **Non-signaling Second Order decide**, but without the modulo operation on the lower order signal $\mathbf{s_l}$. The second order signaling agent is different than the zero and first order signaling agents in the sense that it uses order beliefs $\mathbf{b_o}$ rather than tendencies $\mathbf{t}$, as this and the other second order agents in the study aim to understand which of the lower orders is more dominant in the population and act according to that. The order beliefs allow the second order agent to act as if it were a first order agent, while also keeping the overall consistency of second order agents using order beliefs throughout the study as their results will be compared down the line.

2.2.2 Receiving Agents

The receiving agents make use of two different sets of beliefs, one 1-dimensional and one 2-dimensional. The 1-dimensional beliefs array of the shape (23,1) is used to process the incoming signals, and the 2-dimensional beliefs array of the shape (23,23) is used to decide on an action according to the current beliefs formulated through received signals. How each order of ToM deals with the processing the signal, decision and update processes are explained in the following subsections.

2.2.2.1 Zero Order Receiving Agent

The zero order receiving agent employs a simple greedy strategy for processing the signal and deciding. In the **process_signal** process, the agent uses the the belief vector $\mathbf{b_s}$, where:

$$\mathbf{b_s} = [b_0, b_1, \ldots, b_{22}]$$

on this vector, which is re-initialised every round, the agent performs the same **update** process as the Non-signaling Zero Order Agent, with the only difference being instead of $\mathbf{b}$, the vector $\mathbf{b_s}$ is used. The vector is re-initialised because it represents the agent's beliefs about the most dominant signal in the current round, and the agent needs to re-evaluate this belief every round. The learning rate for the update operations of this vector is 4 times the simulation-wide learning rate $LEARNING\_SPEED = 0.1$. For the decide and update processes the connected beliefs vector $\mathbf{b_c}$ is used, where:

$$\mathbf{b_c} = [[b_{c_{0_0}}, b_{c_{0_1}}, \ldots, b_{c_{0_{22}}}], \ldots, [b_{c_{22_0}}, b_{c_{22_1}}, \ldots, b_{c_{22_{22}}}]]$$

The **decide** process is given by, where ψ represents the zero order signal belief for the round:

$$\psi = \arg\max_s b_{s_i}, \quad i = 0, 1, \ldots, 22.$$

$$d_0 = \arg\max_d b_{c_{\psi_d}}, \quad d = 0, 1, \ldots, 22.$$

The **update** process is the same as the Non-signaling Zero Order Agent, with the difference being that the connected beliefs vector $\mathbf{b_{c_\psi}}$ is used instead of $\mathbf{b}$, and the following calculation is used:

$$b_{c_{\psi_{a_j}}} = b_{c_{\psi_{a_j}}} + LEARNING\_SPEED$$

where a_j represents an action from the actions list for that round.

2.2.2.2 First Order Receiving Agent

The first order receiving agent is a lot like the non-signalling first order agent, in the sense that it simply uses a model of a zero order agent to operate, with the sole addition of having a **process_signal** process, where the **process_signal** process of the zero order agent is called. This agent is essentially a superset of the non-signaling first order agent, with the extra ability to process signals.

2.2.2.3 Second Order Receiving Agent

The second order receiving agent is also almost identical the non-signaling second order agent, with only the extra **process_signal** process, in which the **process_signal** process of the zero and first order agents are called. This agent is also essentially a superset of the non-signaling second order agent, with the extra ability to process signals.

2.2.3 Simulation

The signaling simulation will have a total of 11 different initial population configurations to be simulated. These population configurations are chosen specifically because for each ToM order it again investigates the effects listed for the signaling simulation, but also the effects of scenarios in which there are more signaling agents than receiving agents,

and vice versa, based on the default order population of the simulation (i.e. the first row of Table 2.1). The ratio of difference for the scenarios with more signaling agents than receiving agents is 30:70 and for the scenarios with more receiving agents it's 70:30. The population configurations are like described below:

- Order-wise, the initial 7 scenarios from Table 2.1 are exactly the same, with each order being divided into half for signaling and receiving agents.
- The other 4 scenarios are with the order configuration of $\{33, 33, 33\}$ but with the signaling agents being 70% and 90% of the population and the receiving agents being 30% and 10% of the population, and vice versa.

Throughout the simulation, the behavior of the agents will be observed and recorded. This data will include the signals sent by the signaling agents, the actions chosen by all agents, and the payoffs received by each agent. The data will be analyzed, for each simulation individually but also inter-simulation, to determine how the inclusion of signaling in the Mod Game affects the behavior of Theory of Mind agents. The results will be compared to the findings of De Weerd et al. (2013) and Veltman et al. (2019) to evaluate the effect on signaling on emergent agent behaviour.

3 Results

The results section will be divided into two main subsections: the non-signaling simulation results and the signaling simulation results. Each subsection will contain the results of the simulation for each of the initial population configurations, and will be analyzed individually and inter-simulation in the discussion section to determine the effect of signaling and different order configurations on the behavior of Theory of Mind agents in the Mod Game.

All the initial population configurations of the non-signaling simulation were ran on a total population of 300 agents. The simulation was ran for a total of 50 times for each initial population configuration to get mean results over 50 runs, with each being ran for 1000 epochs. To assess the results, we will use the metric Average Order Agent Score (AOAS), which is the average score of agents for a particular order, averaged over all the 50 runs:

$$AOAS = \frac{1}{50} \sum_{i=1}^{50} \frac{1}{n} \sum_{j=1}^{n} s_{ij} \tag{3.1}$$

where s_{ij} is the score (over 1000 epochs) of the jth agent in the ith run, and n is the total number of agents of that order. The AOAS is used to determine the average score of agents of a particular order, and is used as a uniform metric to compare the results of the different initial population configurations. Another metric that will be used is the standard deviation of each order averaged over the 50 runs, which will be used to determine the variance of the results of each order. Finally, we will use the score difference between orders to determine the difference in performance between orders and how equally the scores are distributed between orders, as a metric to try to measure cooperation. The score difference is provided with the notation $(\mathbf{sd_1^0}, \mathbf{sd_2^1})$, where $\mathbf{sd_1^0}$ is the score difference between the zero and first order agents, and $\mathbf{sd_2^1}$ is the score difference between the first and second order agents. The score difference is simply calculated by subtracting the AOAS of the lower order from the higher order.

3.1 Non-Signaling Simulation

3.1.1 Default Equal Population

The default equal population consisted of 100 agents for each order, with total equality between populations with different orders of Theory of Mind. The end results are shown in Table 3.1, in which the AOAS is shown for each order, with their respective standard deviations.

Order	AOAS	Std Dev
0	2.475	0.341
1	83.448	0.814
2	103.422	18.307

Table 3.1: The average order agent scores for each order in the the default equal population

The results show that the zero order agents, due to them having no Theory of Mind, have the lowest AOAS, **2.475**, with a fairly low standard devia-

tion. The first order agents have the second highest AOAS, **83.448**, with even lower standard deviation considering the higher score. The second order agents have the highest AOAS, **103.422**, but with also the highest standard deviation, which will be explored further as the distribution of the second order scores will be plotted. These results indicate that in the current scenario with the default equal population, the second order agents are the most successful, followed by the first order agents, and then the zero order agents. The score difference between orders is $(\mathbf{sd_1^0 = 80.973}, \mathbf{sd_2^1 = 19.974})$, which indicates that the competitive advantage of the second order agents over the first order agents is not as high as the competitive advantage of the first order agents over the zero order agents. This score difference will also be used as baseline for comparison with the other initial population configurations. Because of the high standard deviation for second order agents, we take a closer look at the distribution of scores of these agents, shown in the plot below:

The distribution plot above shows that there is a bimodal distribution instead of a normal distribution. Considering that this is the result of 50 runs of 1000 epochs, the second order agents seem to be persistent in their order beliefs of whether there are more zero order or first order agents in the population. Due to this bimodal distribution, the standard deviation is high. It is also worthy to note that the group with the lower score has near-identical scores with the first order agents, suggesting that those are the ones that act with the belief there are more zero order agents than first order in the population. To further investigate whether there was a significant difference between the first and second order agent scores, an independent t-test was conducted between the first and second order results, as the large number of iterations validates the use of it through the Central Limit Theorem, for this and further settings (Kwak & Kim, 2017). The resulting p-value was less than 0.001, suggesting that there is significant difference between the distribution of two results. A negative t-statistic (-77.14) was also obtained, meaning that the mean of the first distribution is lower than the mean of the second distribution, further supporting the previous assertion that the lower group of scores are the first order thinkers.

3.1.2 Zero Order Over and Under Abundance

The zero order over-abundance scenario (percentages shown in the second row of Table 2.1) consisted of 150 zero order agents and 75 first and second order agents each. The zero order under-abundance scenario (percentages shown in the third row of Table 2.1) consisted of 60 zero order agents and 120 first and second order agents each. The end results are shown in Table 3.2, which has two sub-tables in the same format as Table 3.1.

Over-abundance $(150_0/75_1/75_2)$		
Order	AOAS	Std Dev
0	2.453	0.352
1	124.034	1.250
2	109.063	13.018

Under-abundance $(60_0/120_1/120_2)$		
Order	AOAS	Std Dev
0	2.473	0.378
1	51.096	0.531
2	99.207	43.861

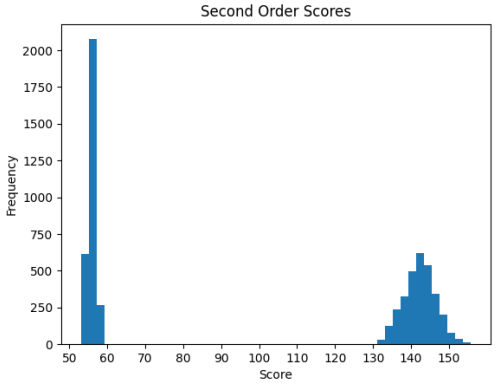
Table 3.2: The average order agent scores for each order in the over-abundance and under-abundance scenarios, where the population is in the format $x_0/y_1/z_2$ for zero, first, and second order agents.

The results show that when there are more zero order agents, the first order agents have more agents to gain scores from and they do. The score difference between orders is $(\mathbf{sd_1^0 = 121.581}, \mathbf{sd_2^1 = -14.971})$, which indicates a drastic advantage shift to first order agents. Com-

pared to the default equal population score difference, the differences between the first order and the zero order increased by **40.608**, while the difference between the second order and the first order decreased by **33.945**. This is in line with our expectations, as the first order agents are more efficient at gaining scores from zero order agents than second order agents. The score distribution of the second order agents in this scenario are plotted below for further inspection:

Here we can see a similar phenomenon, where there is a bimodal distribution and the second order agents are persistent in their order beliefs. In this scenario, the group with the higher score has near-identical scores with the first order agents, suggesting that they decided to employ the first order decision making instead of second order decision making. Because there are more zero order agents, this line of reasoning that the first order thinking would earn more scores is expected and shown in the plot. The results of a t-test show that first order agents obtain significantly higher scores than second order agents ($t = 62.77, p < 0.001$)

When there are fewer zero order agents, the first order agents have fewer agents to gain scores from and they do here too. The score difference between orders is ($\mathbf{sd_1^0 = 48.623}, \mathbf{sd_2^1 = 48.111}$), which indicates an even more advantageous scenario for second order agents. Compared to the default equal population score difference, the differences between the first order and the zero order decreased by **32.350**, while the difference between the second order and the first order increased by **28.137**. The standard deviation for the second order agents also increased, which will be explored again in a distribution plot below:



In this plot, we see the most distinct two modes of the bimodal distribution in the regular simulation, which also correlates to the highest standard deviation observed. The group with the lower scores is near-identical with the first order scores, which would suggest that they are employing the first order reasoning and the other group, second order reasoning. The t-test results also agree, with once again a p-value less than 0.001 and a negative t-statistic (-86.32), suggesting lower score for first order agents and thinking.

3.1.3 First order Over and Under Abundance

The first order over and under-abundance scenarios are given the population configurations $75_0/150_1/75_2$ and $120_0/60_1/120_2$. The results are shown in Table 3.3, in the same format as Table 3.2.

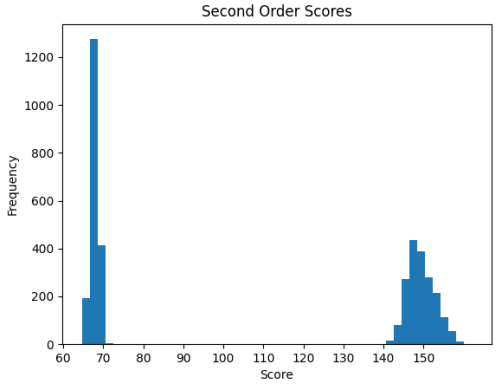
Over-abundance ($75_0/150_1/75_2$)		
Order	AOAS	Std Dev
0	2.473	0.379
1	63.203	0.650
2	107.780	41.247

Under-abundance ($120_0/60_1/120_2$)		
Order	AOAS	Std Dev
0	2.458	0.341
1	99.705	0.990
2	101.938	3.497

Table 3.3: The average order agent scores for each order in the over-abundance and under-abundance scenarios for the first order.

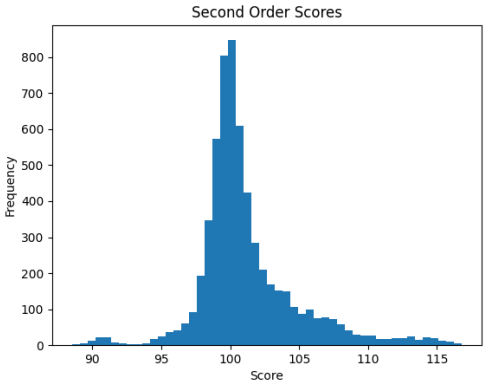
The results show that when there are more first

order agents, the second order agents have more "prey" due to their ability to gain scores from first order agents. The score difference between orders is $(\mathbf{sd_1^0 = 60.730}, \mathbf{sd_2^1 = 44.577})$, which indicates a noticeable advantage shift to second order agents. Compared to the default equal population score difference, the differences between the first order and the zero order decreased by **20.243**, while the difference between the second order and the first order increased by **24.604**. This reflects the decreased opportunity for first order agents to gain scores due to fewer number of zero order agents and the increased opportunity for second order agents to gain scores due to more number of first order agents, which are more deterministic in their behaviour than the zero order agents (standard deviation percentage relative to the mean is **13.87**% for the zero order agents and **0.99**% for the first order agents). The relatively high standard deviation in the second order agent scores is explored in the distribution plot below:



This distribution plot, just like the one below, highlights a stark difference between the scores of the first order thinkers and second order thinkers among the second order agents. The lower scores once again are in-line with the first order scores, which would suggest the higher scores are the result of second order thinking. The negative t-statistic (-96.91) and less than 0.001 p-value once again support that and the significance of the distribution between results.

When there are fewer first order agents, there are more zero order agents, which reflects the trend change favoring the first order, as shown in the Table 3.2 for the zero order over-abundance scenario, but in a less drastic manner. The score difference between orders is $(\mathbf{sd_1^0 = 97.247}, \mathbf{sd_2^1 = 2.233})$, which is almost a mean of the score difference for the default equal population $(\mathbf{sd_1^0 = 80.973}, \mathbf{sd_2^1 = 19.974})$ and the zero order over-abundance scenario $(\mathbf{sd_1^0 = 121.581}, \mathbf{sd_2^1 = -14.971})$. This shows that the trend is stable and more zero order agents mean more scores for the first order agents. In this case we have the lowest standard deviation across orders so far, with the second order agents having a notable coefficient of variation of **3.43**%. To further investigate, below is their score distribution plot:



This distribution plot shows a normal distribution, unlike the others. When the first order agents are under-abundant, the second order agents must have all perceived the zero order to be more dominant and just acted with the first order model, with more variation due to the extra epsilon check of course. The mean scores of the first and second order being very close, within the window of the standard deviation, also suggests that. Interestingly, however, the score of the second order agents is significantly higher than the score of first order agents $(t = -17, p < 0.001)$ which could be caused by the extra epsilon check.

3.1.4 Second Order Over and Under Abundance

The second order over and under-abundance scenarios are in the same format as the zero and first order over and under-abundance scenarios, but with the second order agents. The end results are shown in Table 3.4.

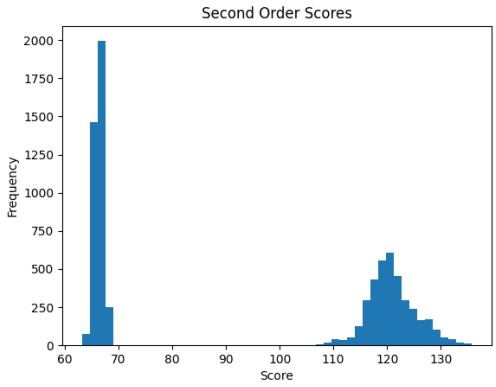
The results show that when there are more second order agents, the overall score decreases with the first order agent scores tak-

Over-abundance ($75_0/75_1/150_2$)		
Order	AOAS	Std Dev
0	2.477	0.361
1	63.268	0.582
2	93.250	27.571

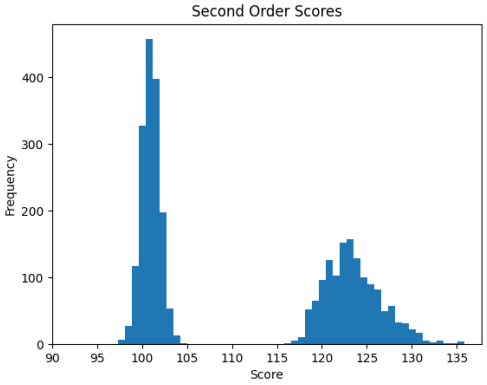
Under-abundance ($120_0/120_1/60_2$)		
Order	AOAS	Std Dev
0	2.482	0.356
1	99.689	0.979
2	111.518	11.331

Table 3.4: The average order agent scores for each order in the over-abundance and under-abundance scenarios for the second order.

ing a decent hit. The score difference between orders is $(\mathbf{sd_1^0 = 60.791}, \mathbf{sd_2^1 = 29.982})$. When there are fewer second order agents, both the first and second order agents have higher scores. The score difference between orders is $(\mathbf{sd_1^0 = 97.207}, \mathbf{sd_2^1 = 11.829})$. Similar to the previous findings, second order agents obtain a higher score than first order agents both in the over-abundance ($t = -67.43, p < 0.001$) and the under-abundance ($t = -81.03, p < 0.001$) scenarios. The score distribution for both scenarios are plotted below, with the over-abundance being first and under-abundance being second:



Here once again a bimodal distribution is seen, with the lower scores correlating to the first order thinking. Even though the zero order / first order ratio is the same as the default equal population, the standard deviation is slightly higher, suggesting a more divided order beliefs in the second order population due to less number of other agents (150 total instead of 200) to base their beliefs on.



This plot also has the same bimodal distribution scenario, where the lower group of second order scores correlate with first order scores, thus first order thinking, and the higher group of scores correlating to second order thinking.

3.2 Signaling Simulation

For the tables presented in this subsection, to denote the inital population, a similar format of:

$$x_{s_0}/y_{s_1}/z_{s_2}/a_{r_0}/b_{r_1}/c_{r_2} \tag{3.2}$$

is used, where x, y, and z are the number of 0, 1, and 2 order signaling agents, and a, b, and c are the number of 0, 1, and 2 order receiving agents respectively. The s and r subscripts denote signaling and receiving agents in an understandable format.

3.2.1 Default Equal Population

The default equal population consisted of 100 agents of each order, with each order being split into half signaling and half receiving agents. The end results are shown in Table 3.5, where the signaling agents are on the left and the receiving agents are on the right and are labelled as such.

The overall order-wise results, which groups signaling and receiving agents together, are shown in Table 3.6.

The results in Table 3.5 offer numerous insights and base conclusions that can be used in future ones. In terms of zero order agents, they have almost identical scores, which means their respective higher order agents either have no effect or coincidentally the same effect on their scores. This is remarkable considering the signalers base their decision on their own signal while the receivers base

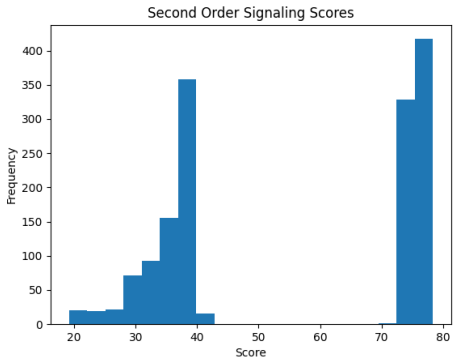
$50_{s_0}/50_{s_1}/50_{s_2}/50_{r_0}/50_{r_1}/50_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	4.525	0.252	4.550	0.245
1	13.896	17.429	78.209	1.043
2	58.253	24.377	69.144	6.549

Table 3.5: The average order agent scores for each order in the default equal population for the signaling simulation.

Order	AOAS	Std Dev
0	4.538	0.248
1	46.053	9.236
2	63.699	15.463

Table 3.6: The average order agent scores for each order in the default equal population for the signaling simulation.

their decision on the "dominant" signal. In terms of first order agents, the signaling agents have a much lower score than the receiving agents, which have the leading score across all order/agent type combinations with 78.209 AOAS. First order signaling agents also have high standard deviations, which means their behaviour is not very deterministic and makes sense given that they base their decision on their own signal. In terms of second order agents, the receiving agents have a higher score than the signaling agents, but the difference is not as large as the first order agents. Second order receivers have a lower score than first order receivers, given the high variance of first order senders making their behaviour harder to predict by second order receivers. Table 3.6 tells us a different story, where compared to the scores of the non-signaling default equal population in Table 3.1, there is a more equal distribution of the scores across orders and the overall average order agent score is lower. The score difference between the orders is ($\mathbf{sd_1^0 = 41.515, sd_2^1 = 17.646}$). The distributions of the second order signaling and receiving agents were plotted below, to see if the high standard deviation in second order signaling agents is due to a similar bimodal distribution like the regular simulation, and if the relatively lower standard variation in second order receiving agents is due to a unimodal distribution:



The distribution plots show that the second order signaling agents, whose workings are very similar to the regular second order agents, follow a similar bimodal distribution where the order beliefs separate the agents into two groups that act persistently on either a first way of thinking or a second way of thinking. The distribution plot for the second order receivers are different than the rest, as the distribution looks to have two peaks, suggesting a bimodal distribution, however the peaks aren't separate like the rest of the distribution plots presented so far. This suggests that second order agents may be shifting their order beliefs instead of being persistent in them. The same trend of distributions were observed within other scenarios, where the signaling distribution was bimodal and separate, and the receiving distribution was bimodal but not separate and uniform. Also a likely reason may be the variation of the first order signalers making it harder for second order receivers to obtain consistently high scores.

3.2.2 Zero Order Over and Under Abundance

The zero order over and under-abundance scenarios are the same as the zero order over and under-abundance scenarios in the non-signaling simulation, with each order being split into half signaling and half receiving agents. The end results are shown in Table 3.7.

$75_{s_0}/37_{s_1}/37_{s_2}/75_{r_0}/38_{r_1}/38_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	4.026	0.293	4.039	0.279
1	15.140	21.570	86.602	1.175
2	55.402	34.814	88.039	3.657
$30_{s_0}/60_{s_1}/60_{s_2}/30_{r_0}/60_{r_1}/60_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	5.143	0.252	5.127	0.227
1	11.889	13.713	71.605	0.939
2	58.529	17.660	59.255	6.802

Table 3.7: The average order agent scores for each order in the zero order over and under-abundance scenarios for the signaling simulation

The overall order-wise results, which groups signaling and receiving agents together, are shown in Table 3.8.

$75_{s_0}/37_{s_1}/37_{s_2}/75_{r_0}/38_{r_1}/38_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	4.033	0.286
1	50.871	11.372
2	71.721	18.921
$30_{s_0}/60_{s_1}/60_{s_2}/30_{r_0}/60_{r_1}/60_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	5.135	0.240
1	41.747	7.326
2	58.892	12.231

Table 3.8: The average order agent scores for each order in the zero order over and under-abundance scenarios for the signaling simulation

The results show that when there are more zero order agents, the total average order agent score increases with only the zero order agents experiencing a loss in AOAS. This is because the more zero order agents there are, the more ways of earning score for first and second order agents. The score difference between orders is $(\mathbf{sd_1^0 = 46.838}, \mathbf{sd_2^1 = 20.85})$ which shows the overall score increase compared to the default equal population. A notable performance increase can be seen in second order receiving agents, which surpass the first order signaling agents in AOAS.

When there are fewer zero order agents, the total average order agent score decreases with only the zero order agents experiencing a gain in AOAS. This is in line with the over-abundance results as with fewer zero order agents the opposite order-wise effects are seen. The score difference between orders is $(\mathbf{sd_1^0 = 36.612}, \mathbf{sd_2^1 = 17.145})$ which shows the overall score decrease compared to the default equal population.

3.2.3 First Order Over and Under Abundance

The first order over and under-abundance scenarios are the same as the first order over and under-abundance scenarios in the non-signaling simulation, with each order being split into half signaling and half receiving agents. The end results are shown in Table 3.9.

$37_{s_0}/75_{s_1}/37_{s_2}/38_{r_0}/75_{r_1}/38_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	3.991	0.269	3.972	0.263
1	13.860	18.372	86.269	1.062
2	57.635	32.302	70.885	10.415
$60_{s_0}/30_{s_1}/60_{s_2}/60_{r_0}/30_{r_1}/60_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	4.526	0.269	4.532	0.244
1	13.275	17.989	71.823	0.901
2	49.782	24.620	78.312	2.252

Table 3.9: The average order agent scores for each order in the first order over and under-abundance scenarios for the signaling simulation

The overall order-wise results, which groups signaling and receiving agents together, are shown in Table 3.10.

$37_{s_0}/75_{s_1}/37_{s_2}/38_{r_0}/75_{r_1}/38_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	3.982	0.266
1	50.064	9.717
2	64.260	13.426
$60_{s_0}/30_{s_1}/60_{s_2}/60_{r_0}/30_{r_1}/60_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	4.529	0.256
1	42.549	9.445
2	64.047	13.436

Table 3.10: The average order agent scores for each order in the first order over and under-abundance scenarios for the signaling simulation

The results show that when there are more first order agents, the zero order score decreases and the first order score increases, with the second order score remaining relatively constant. The score difference between orders is $(\mathbf{sd_1^0 = 46.082}, \mathbf{sd_2^1 = 14.196})$. In this configuration, the only group with more score compared to the default equal population is first order receiving agents, which prove time and time again that they are the most advantageous ones in the signaling simulation in most cases.

When there are fewer first order agents, the zero and first order scores decrease a bit, with the overall scores remaining relatively constant. The score difference between orders is $(\mathbf{sd_1^0 = 38.020}, \mathbf{sd_2^1 = 21.498})$. The behaviour of second order agents is particularly peculiar, as the signaling second order agents have lower score but the receiving second order agents have higher score. This could suggest that while the second order signaling agents can mainly benefit from first order agents, the second order receiving agents can benefit from both first and second order agents, signaling or receiving.

3.2.4 Second Order Over and Under Abundance

The second order over and under-abundance scenarios are the same as the second order over and under-abundance scenarios in the non-signaling simulation, with each order being split into half signaling and half receiving agents. The end results are shown in Table 3.11.

$37_{s_0}/37_{s_1}/75_{s_2}/38_{r_0}/38_{r_1}/75_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	5.269	0.212	5.301	0.210
1	14.202	16.219	60.157	0.774
2	53.502	11.432	64.183	1.449
$60_{s_0}/60_{s_1}/30_{s_2}/60_{r_0}/60_{r_1}/30_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	4.015	0.301	4.003	0.293
1	12.944	19.765	92.317	1.313
2	59.171	37.668	83.654	9.141

Table 3.11: The average order agent scores for each order in the second order over and under-abundance scenarios for the signaling simulation

The overall order-wise results, which groups signaling and receiving agents together, are shown in Table 3.12.

$37_{s_0}/37_{s_1}/75_{s_2}/38_{r_0}/38_{r_1}/75_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	5.285	0.211
1	37.179	8.497
2	58.843	6.441
$60_{s_0}/60_{s_1}/30_{s_2}/60_{r_0}/60_{r_1}/30_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	4.009	0.297
1	52.630	10.539
2	71.413	23.404

Table 3.12: The average order agent scores for each order in the second order over and under-abundance scenarios for the signaling simulation

The results show that when there are more second order agents, the zero order AOAS increases and the first and second order AOAS decreases, with the first order signaling agents having a small increase in score. The score difference between orders is $(\mathbf{sd_1^0 = 31.194}, \mathbf{sd_2^1 = 21.664})$. One notable phenomenon observed in this scenario is the noticeably lower standard deviation in second order overall AOAS, which is lower than the value of 10 only

in this scenario.

When there are fewer second order agents, the overall total AOAS jumps from **114.29** in the default equal population to **128.052**. Other than the zero order agents, the other orders all have more overall AOAS, with the first order receiving agents once again being on the top with **92.317** AOAS. The score difference between orders is $(\mathbf{sd_1^0 = 48.621}, \mathbf{sd_2^1 = 18.783})$. The standard deviation of the second order overall AOAS is also the highest so far in this scenario, at **23.404**.

3.2.5 Signaling and Receiving Over-Abundance

In the signaling and receiving over-abundance scenarios, the orders have equal distribution of agents, like the default equal population scenario. The difference is that there is a 70/30 ratio between signaling and receiving agents, and the inverse. The population percentages are $70_{s_0}/70_{s_1}/70_{s_2}/30_{r_0}/30_{r_1}/30_{r_2}$ and $30_{s_0}/30_{s_1}/30_{s_2}/70_{r_0}/70_{r_1}/70_{r_2}$. The end results are shown in Table 3.13.

$70_{s_0}/70_{s_1}/70_{s_2}/30_{r_0}/30_{r_1}/30_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	5.125	0.275	5.123	0.284
1	11.854	16.420	107.980	1.433
2	66.860	45.308	73.014	27.055

$30_{s_0}/30_{s_1}/30_{s_2}/70_{r_0}/70_{r_1}/70_{r_2}$				
	Signaling		Receiving	
Order	AOAS	Std Dev	AOAS	Std Dev
0	3.761	0.263	3.766	0.269
1	13.943	18.462	48.878	0.636
2	42.050	9.681	79.185	17.140

Table 3.13: The average order agent scores for each order in the singaling and receiving over-abundance scenarios for the signaling simulation

The overall order-wise results, which groups signaling and receiving agents together, are shown in Table 3.14.

The results show that when there are more signaling agents, every group of agents have an increased AOAS barring the exception of the first order signaling agents, which have a lower score than the default equal population. The score difference between orders is $(\mathbf{sd_1^0 = 54.793}, \mathbf{sd_2^1 = 10.020})$. Notable phenomena are the first order receiving agents' AOS being highest in this and all scenarios with a score of **107.980**, and the second order agents having a relatively high standard deviation in their AOAS, at **45.308** for signaling agents and **27.055** for the receiving agents respectively.

When there are more receiving agents, all agent groups excluding the second order receiving agents and first order receiving agents see a drastic decrease in their scores, with the second order agents seeing a relative increase. The score difference between orders is $(\mathbf{sd_1^0 = 27.646}, \mathbf{sd_2^1 = 29.208})$. The second order receiving agents performing relatively good in both scenarios is notable, and might suggest that they will perform relatively well in any given scenario.

$70_{s_0}/70_{s_1}/70_{s_2}/30_{r_0}/30_{r_1}/30_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	5.124	0.279
1	59.917	8.926
2	69.937	36.182

$30_{s_0}/30_{s_1}/30_{s_2}/70_{r_0}/70_{r_1}/70_{r_2}$		
	Overall	
Order	AOAS	Std Dev
0	3.764	0.266
1	31.410	9.549
2	60.618	13.411

Table 3.14: The average order agent scores for each order in the signaling and receiving over-abundance scenarios for the signaling simulation

4 Discussion

Before getting into the discussion, it would be beneficial to give a reader's note regarding the terminology. As outlined in Methodology, the mathematical formulas the agents use for their decision, signaling and updating processes are meant to get the most favourable outcome for that individual agent in particular. Essentially, they are "greedy" algorithms that intend to maximise the agent score. In this study, we are more focused on the outcome and the score distribution of agents, and hope to measure how drastic the difference between scores

of different orders of ToM agents to determine how "fairly distributed" the end results are. In this and the following section, we will refer to strategies as "cooperative" when they end up with more fairly distributed end results with less score difference between orders compared to the default equal population, while we will refer to strategies as "greedy" when they result in more skewed score distributions with similar or more score difference between the orders compared to the default equal population.

4.1 The Regular Simulation

In the regular simulation, it is seen that including the default equal population, in 6 out of 7 tested scenarios, agents perform better with a higher order ToM with the first order agents performing better than zero order agents and second order agents performing better than the first order agents.

The only exception is the $150_0/75_1/75_2$ scenario, in which there are 50% more zero order agents than the default equal population. In this scenario, the first order agents perform better than the second order agents but also this is the scenario with the highest total score across all three orders of ToM. This is because while second order agents can adapt to a zero order dominant population, first order agents are ready for it and benefit from it from the beginning. This is evident in all the other scenarios too, as the first order score is directy correlated with the number of zero order agents.

The second order agents are not as sensitive to population changes, but they also prefer more zero order agents than more first order agents, with them having higher AOAS in scenarios with more zero order agents.

These results indicate that without signaling, unless there is a significant (around 50%) increase in the number of zero order agents, higher order agents, in general, perform better than lower order agents (with the highest order being 2). The result of the several independent t-tests conducted across scenarios also support the significant difference between the first and second order results. The high standard deviation seen in second order agents are caused by the bimodal distribution of the second order agents, caused by their order beliefs. The average agent score (across all orders) for 300 agents and 1000 epochs is **63.115** for the default equal population, and **78.517** for the $150_0/75_1/75_2$ scenario, which has the highest average.

4.2 The Signaling Simulation

In the signaling simulation, the picture is not too different from the regular simulation if analysed from an order-wise perspective. In all of the scenarios, the first order agents collectively perform better than zero order agents, and second order agents collectively perform better than first order agents, with no exceptions. This is expected, as also seen in the regular simulation and in previous work, higher order agents perform better than lower order agents, with the highest order being the second.

The main difference between the regular and signaling simulations is the score differences. In the regular simulation, in all scenarios, at least one of the score differences (either $\mathbf{sd_1^0}$ or $\mathbf{sd_2^1}$) is higher than **48** whereas in the signaling simulation, in 8 out of 9 scenarios, both score differences are lower than **48**. The only exception is the $60_{s_0}/60_{s_1}/30_{s_2}/60_{r_0}/60_{r_1}/30_{r_2}$ scenario where $\mathbf{sd_1^0 = 48.621}$. This indicates a more balanced distribution of scores between the orders of ToM in the signaling simulation, from a higher, order-wise perspective.

The average agent score (across all orders) for 300 agents and 1000 epochs is **38.197** for the default equal population, and **43.684** for the $60_{s_0}/60_{s_1}/30_{s_2}/60_{r_0}/60_{r_1}/30_{r_2}$ scenario, which has the highest average. Both of these scores are lower than their counterparts in the regular simulation, which is expected as the signaling simulation is a more difficult task for the agents that contains more than one dimension of learning. While this may also be attributed to the fact that the signaling simulation is simply detrimental to the agents, the fact that the score distribution between the orders of ToM is more balanced in the signaling simulation indicates that the resulting environment and the behavior caused by that environment is not "worse" than the regular simulation, but rather different and more complex.

In fact, when the results of the signaling and receiving agents are compared in-order, it is seen that the receiving agents perform better than the signaling agents in all scenarios, except for zero order agents, whose behaviour is not as affected by the signaling/receiving distinction. Before the

experimentation process started, we hypothesized that the signaling agents would learn how the receiving agents respond to their signals and would adapt to that in a greedy manner, and would earn higher scores than the receiving agents. However, the results show that the receiving agents perform better than the signaling agents, which indicates that the receiving agents are able to adapt to the signaling agents' signals and respond to them in a way that is more beneficial to them than the signaling agents. This phenomenon can be explained by the processing order of the agents, as the receiving agents process the signals after the signaling agents signal them, in which case a simulation with asynchronous processing could be the topic of future work. If the same results are observed in an asynchronous simulation, then it would be safe to say that the receiving agents are more advantageous than the signaling agents in the signaling simulation. Another possible contributing reason is that while receiving agents process all the signals transmitted and update accordingly, the signaling agents only take their signal into consideration. This again could be addressed in future work along with the zero order signaling agent update process, which could have also contributed to this phenomenon.

Inspecting the results further, one can see that in 6 out of 9 scenarios, the first order receiving agents perform the best, with the second order receiving agents performing the best in the remaining 3 scenarios. This is quite interesting, as both in the regular and signaling simulation the second order agents have better overall scores than first order agents. This could suggest that for receiving agents, the "higher order ToM advantage" starts and ends at the first order, and the scores for even higher orders of ToM than tested (3, 4, etc.) would be lower than the scores of the previous order. This, once again could be explored in future work.

The differences between distributions of the second order signaling and receiving agents should be noted, as the signaling agents score distribution is very similar to the second order score distribution in the regular simulation, with a bimodal distribution comprised of two separate groups of results with far peaks. In the second order receiving agents score distribution however, a not separate, more uniform distribution is observed, while still bimodal, with two close peaks.

Like in the regular simulation, more zero order agents mean more score for the other orders, and the reverse, but this time, the second order agents are more sensitive to the number of zero order agents than the first order agents. This is evident in the difference between second order scores in the default equal population and the $75_{s_0}/37_{s_1}/37_{s_2}/75_{r_0}/38_{r_1}/38_{r_2}$ scenario, which is higher than the difference in first order agent scores. An increased number of first order agents benefits only the first order agents and does not have visible significant effect on other order scores while the opposite situation has the opposite effect on first order agents and once again, no visible significant effect on other order scores. Looking at the second order over/under-abundance scenarios, it can be seen that the second order agents decrease all order agent scores when there's more of them, and increase all order agent scores when there's less of them. This is the opposite of the first order agents, which increase all order agent scores when there's more of them, and decrease all order agent scores when there's less of them. These last two claims exclude the zero order agents, as they are not as affected by the number of other order agents as the other orders are. This is also evident in the regular simulation, where the zero order agents are not as affected by the number of other order agents as the other orders are.

The situation with the signaling/receiving over-abundance scenarios is quite peculiar. An over-abundance of signaling agents benefit every agent group except first order signaling agents which have a lower score, while and ober-abundance of signaling agents benefits only the first order signaling and second order receiving agents, with the rest having lower scores than the default equal population. This suggests three main things:

- Our previous suggestion about receiving agents employing a greedy strategy hold up, as more of them means worse or equal scores for all agent groups except second order receiving agents.
- More signaling agents means better scores for all agent groups except first order signaling agents, which suggests that they might not employ a greedy strategy but rather a cooperative one, which is beneficial to all agent groups.
- First order signaling agent behaviour is unex-

pected and should be explored further, could maybe have been caused by wrong formulation or implementation of the agents.

Overall, when we compare the signaling simulation results with the regular simulation results, we can see that the signaling simulation is a more complex task for the agents, as the overall average AOAS lower in the signaling simulation than in the regular simulation. However, the score distribution between the orders of ToM is more balanced in the signaling simulation, which indicates that the resulting environment and the behavior caused by that environment is not "worse" (in terms of cooperation and collective gains, which are indicated by the score differences) than the regular simulation, but rather different and more complex. The results with more even distribution of scores between the orders of ToM could both be attributed to the fact that because it is a more complex task, it is harder to employ a greedy strategy, but also to the fact that the agents are able to employ cooperative or semi-cooperative strategies due to the added dimension of cost-free communication.

5 Conclusion

In this study, we have tried explore the intricate interplay between Theory of Mind (ToM) and strategic decision-making in competitive settings, in a very specific setting we have formulated. Our primary research question centered around the effects of ToM on agent behavior, with a specific focus on our setting: the Mod Game with signaling. By simulating a total of 16 simulations with different agent compositions, we hoped to shed light on the effect of communication in competitive settings like this where there's traditionally no communication. Even though previous studies with the non-signaling version of the Mod Game exist, we developed an in-house simulation for both the signaling and non-signaling simulations inside the same framework for easier comparison and validity of results.

In the regular simulation, it has been observed that the agents with higher ToM agents were usually out-performing lower ToM agents, employing mostly greedy strategies. This was reflected in the result trends, where the score of $\mathbf{N}$ Order ToM agents were more often that not directly correlated with the number of $\mathbf{N-1}$ Order ToM agents. Only in 1 out of 7 simulations the higher order ToM dominance was not observed, with the first order agents performing better than second order agents, which strengthens the hypothesis that higher order ToM agents are more likely to outperform lower order ToM agents, in the absence of communication, the significant difference between the first and second order agent scores were supported with distribution plots and following independent t-tests.

In the signaling simulation, a similar trend is observed order-wise, higher order beats the lower order. However in each order, we had two groups of agents: one signaling and one receiving. When these sub-groups were inspected individually, an unexpected trend was observed, with the first order receiving agents being the highest performers in 5 out of 9 simulations. This was surprising for two reasons: 1. There is a clear trend of lower order advantage over an higher order and 2. we expected the signaling agents to perform better than the receiving agents, because we thought they would employ deceptive strategies with signaling first and deciding second. However, the results showed that the receiving agents were the ones who were advantageous and getting more scores. While this could be attributed to them simply employing a greedy strategy over signaling agents, there are results that say otherwise.

The score differences between orders, which is the metric we used to measure cooperation on an abstract level, is visibly and significanly lower the signaling simulation than the regular one. This metric shows the "inequality" between the score distribution between orders and essentially is a measure of how harsh higher orders of ToM affect the scores in competitive settings such as this. We believe, while it may simply be harder to effectively employ greedy strategies due to the complexity of the simulation compared to the regular one, that the addition of cost-free communication has a significant effect on the scores and the cooperation between agents.

A bimodal distribution with separate peaks was observed in the results for regular second order and second order signaling agents, which suggests that they may be persistent in their order beliefs. In second order receiving agents however, a still bimodal distribution was observed but with close peaks and

no separation between groups of data. This suggests that they may be shifting their order beliefs, due to their different architecture and processes.

Signaling, which is basically cost-free communication on a 1-dimensional level, was added to this scenario to see if agents would be more willing to cooperate given that it would be easier to do so. However, we've uncovered some interesting results that it is not as linear as we have hypothesized beforehand, but leads to way more complex emergent behaviour. The difference in score differences indicate that relatively more cooperative strategies are used, but not clear-cut cooperative or greedy like we have thought. This is an exciting result, as we see clear behavioral changes in agents when communication is introduced, and we believe that this is a very interesting area to explore further. This also supports the claim in Farrell & Rabin (1996) where it is stated that the addition of communication would drastically alter the Nash Equilibria and/or the behaviour of agents in many competitive settings, which we have observed in our simulation. Due to this, we also believe that cost-free communication like this could be introduced to many different competitive settings in future work, to see how it would affect the behaviour of agents in those settings. Especially, the addition of signaling in natural selection-like scenarios could be very interesting to observe, as it could lead to more cooperative strategies being employed by agents, which could lead to more interesting results overall.

One possible limitation of this study is the process order, where the signaling and receiving process is synchronous and in order rather than asycnhronous and free flowing like real life human communication. While it is not a clear limitation per se, as in many human communication scenarios there are rules and a structure, it might be influencing the current behaviour of agents. That's why a follow-up study with asynchronous communication could also be worthwile to see if the results would be different.

Another limitation is the fact that we have only used one type of communication, which is signaling. While it is a very simple form of communication, it is still communication and it could be interesting to see how other forms of communication would affect the results. For example, if we were to introduce a cost to communication, would the results be different? Or if we were to introduce a more complex form of communication, like a 2-dimensional one, would the results be different? These are all interesting questions that could be explored in future work.

In this study, we have simulated the Mod Game with and without cost-free communication played by 300 Theory of Mind agents. We have found out that the addition of communication has a significant effect on the scores and the cooperation between agents, and changes the dynamic and behaviour of agents on a noticeable level. The agents had a more even distribution of scores in the simulation with communication, leading us to believe that the agents employ a relatively more cooperative strategy. While the employed experimental methodology may have been limited in some aspects, we believe that the results merit further exploration of the effects of communication in competitive settings like this.

References

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

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

de Weerd, H., Verbrugge, R., & Verheij, B. (2017). Negotiating with other minds: The role of recursive theory of mind in negotiation with incomplete information. *Autonomous Agents and Multi-Agent Systems*, *31*(2), 250–287.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic perspectives*, *10*(3), 103–118.

Frey, S., & Goldstone, R. L. (2013). Cyclic game dynamics driven by iterated reasoning. *PloS one*, *8*(2), e56416.

Gavrilets, S., & Vose, A. (2006). The dynamics of machiavellian intelligence. *Proceedings of the National Academy of Sciences*, *103*(45), 16823–16828.

Kuhn, S. (2008). Prisoner's dilemma. *Stanford Encyclopedia of Philosophy*.

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, *70*(2), 144–156.

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, *4*(3-4), 231–244.

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Veltman, K., de Weerd, H., & Verbrugge, R. (2019). Training the use of theory of mind using artificial agents. *Journal on Multimodal User Interfaces*, *13*(1), 3–18.

Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and brain sciences*, *11*(2), 233–244.