How Game Structure Affects the Effectiveness of Higher-Order Theory of Mind Reasoning

Daniel Johannes Coertze (d.j.coertze@student.rug.nl)
Faculty of Science and Engineering, University of Groningen
Nijenborgh 9, 9747 AG Groningen, The Netherlands
Supervisor: dr. H.A. (Harmen) de Weerd

July 23, 2023

Abstract

Theory of mind involves the cognitive ability to attribute mental states, such as beliefs, intentions or desires, to oneself and others, acknowledging the potential variations among individuals. This study explores the influence of game structure on the effectiveness of higher-order theory of mind reasoning within a simplified version of Pecking Order, a competitive card game developed by Richard Garfield. The investigation focuses on two distinct game variations, sequential and simultaneous, which shape the overall game structure through differences in information availability, turn order, and round numbers. Agents with Zero-Order and First-Order theory of mind capabilities are introduced to assess the effectiveness of theory of mind reasoning across these variations. Employing simulation theory of mind, these agents consider opponents' mental states and predict their behavior by simulating actions based on their opponents' positions in the game. By analyzing the impact of game structure on the effectiveness of higher order theory of mind reasoning, this research provides valuable insights into how game structure influences the effectiveness of theory of mind reasoning in the context of strategic decision-making games.

1 Introduction

In strategic environments, accurately predicting the behavior of others is crucial for effective decision-making and response strategies. Theory of mind, which involves attributing unobservable mental states to others, offers a promising approach to achieve this goal (Barlassina and Gordon, 2017; de Weerd et al., 2013). This paper explores the advantages of utilizing theory of mind in the context of the card game Pecking Order, where the ability to predict and respond to opponents' moves can significantly impact the outcome of the game.

To investigate the effectiveness of theory of mind in Pecking Order, we employ an agent-based computational model. Our model incorporates Zero-Order and First-Order theory of mind, enabling agents to form beliefs about opponents' behavior and attribute similar belief systems to them in the form of probability distributions. By simulating opponents' mental states and predicting their actions, agents strategically determine which actions they themselves should take to gain a competitive advantage and increase their odds of winning.

In the following sections, we provide a comprehensive analysis of the role and effectiveness of theory of mind in the Pecking Order game. Section 2 outlines the game mechanics and objectives of Pecking Order, providing a foundation for understanding the strategic interactions involved. It then goes on to introduce the agents and their strategies, emphasizing the significance of theory of mind in shaping their decision-making processes. Utilizing our agent-based computational model, we describe the experimental setup in Section 3, including the parameters, simulations, and evaluation criteria used in our study.

The results of our evaluations are presented in Section 4, where we analyze the performance of agents with varying levels of theory of mind. Section 5 offers a discussion of the implications and significance of our findings, considering the broader context of strategic interactions. Lastly, in Section 6, we conclude the paper by summarizing our findings and discussing potential directions for future research.

Our research aims to contribute to the understanding of how theory of mind enhances decision-making processes as a whole by investigating its advantages and limitations within the context of Pecking Order. Through our agent-based computational model, we hope to provide valuable insights into the role and effectiveness of theory of mind in strategic interactions, shedding light on its impact on game outcomes and informing future developments in multi-agent systems.

2 Methods

2.1 Outline of Pecking Order

Pecking Order is a card game in which two players compete to control 'perches' on a board by playing bird-themed cards of varying strengths. The version of Pecking Order used for the purposes of this project is a modified version of the original game by Richard Garfield (1998).

In this modified version, each player starts with 4 unoccupied perches on their side of the board and 4 numbered cards (1-4). The game consists of multiple rounds. In each round, both players play one of their cards face-down on an unoccupied perch on their side of the board. If both players have played a face-down card on the same perch location on opposite sides of the board (either in the current or previous turns), the cards on that perch are flipped over. The player with the higher value card on that perch gains control of it. If equal value cards are played on the same perch, neither player can gain control of it, therefore no points are awarded to either player for that perch. Note that even if a particular player controls a perch, all cards remain on the board until the end of the game.

Controlling a perch earns a player points equal to the perch index (perch 1 = 1 point, perch 2 = 2 points, perch 3 = 3 points, perch 4 = 4 points). A player without control of a perch gains no points.

The game continues until both players have played all their cards and all perches are occupied. At that point, all cards are face-up, and each player calculates their final score by summing the points gained from controlling perches. The player with the most amount of points at the end of the game claims the victory. Note that if both players have attained the same number of points at the end of the game, the player controlling perch 1 will be deemed the winner.

An example of the starting board state in a simplified game of Pecking Order is found in Figure 1. Note that neither player is able to see their opponents hand in an actual game, however each players cards are presented here for visual clarity.

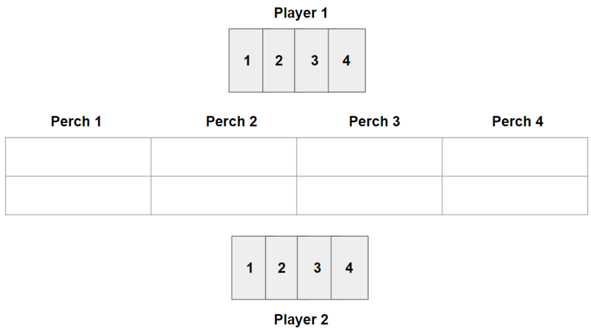

Figure 1: Starting Board State of Pecking Order

This project aims to evaluate how game structure influences the effectiveness of higher order theory of mind. As such, two variations of the modified Pecking Order game are considered. In the sequential variation, players take turns playing their cards in an 'I play, you play' fashion during each round. In the simultaneous variation, both players play their cards simultaneously. These variations affect the amount of information available to each player on their turn and the total number of rounds played. The sequential variation consists of 8 rounds, while the simultaneous variation has 4 rounds.

The project models different types of players (referred to as agents) with varying strategies or levels of theory of mind. This allows for a comparison of the effectiveness of higher order theory of mind reasoning across the two versions of the Pecking Order game. More information about these agents and their strategies is provided in section 2.2.

2.1.1 Similarities to Limited Bidding

This project draws inspiration from a game called Limited Bidding (de Weerd and Verheij, 2011), which, despite being unrelated, shares similarities with Pecking Order. In Limited Bidding, players use numbered tokens instead of cards, and there is no board. Similar to the simultaneous version of Pecking Order, players each play one of their numbered tokens simultaneously during each round. The player with the highest-valued token in a round earns a single point.

Although similarities between the two games exist, Pecking Order and Limited Bidding differ in several aspects. Limited Bidding has perfect information, while Pecking Order has imperfect information. In Limited Bidding, the tokens played by the opponent are immediately revealed on the same turn they are played. This allows for immediate knowledge of the opponent's remaining tokens during each round. In Pecking Order, this is not always the case. Cards played by either player can be revealed in the same round or much later in the game. Additionally, points are accumulated after each round in Limited Bidding, whereas in Pecking Order, they are tallied at the end of the game. Thus, while the game structures are similar, Pecking Order introduces more uncertainty about the opponent's past actions and remaining cards compared to Limited Bidding.

In Limited Bidding, assuming common knowledge of rationality, rational agents play the game randomly (de Weerd and Verheij, 2011). However, experiments with human subjects have shown that humans tend to deviate from complete randomness and exhibit patterns in their reasoning and behavior different to what is predicted by game theory (Bacharach and Stahl, 2000; Fehr and Gächter, 2000; Henrich et al., 2001; McKelvey and Palfrey, 1992; Stahl and Wilson, 1995). When agents play against the same opponent multiple times, they can exploit their opponent's tendencies and adjust their strategy to gain an advantage. From a theory of mind perspective, rational agents in Limited Bidding assume that their opponent also plays rationally, which, in this case, means playing randomly. To prevent providing their opponent with a strategic advantage, a rational agent in Limited Bidding would choose to play randomly each turn, assuming their opponent would do the same.

Furthermore, in Limited Bidding, the absence of a pure-strategy Nash Equilibrium has been demonstrated (de Weerd and Verheij, 2011). In game theory, a Nash Equilibrium occurs when an agent has no incentive to change its strategy given its opponent's strategy. In other words, it is a state where neither player can gain an advantage by unilaterally changing their strategy based on their opponent's current strategy.

In the context of Pecking Order, suppose a pure Nash Equilibrium strategy σ exists. In a game where both players play σ, exactly one will win or both will draw. In this case, the losing player or the players suspecting a draw will have an incentive to play randomly and at least attempt at gaining a chance to win. Thus, σ is not a pure Nash Equilibrium strategy for the loser or the players suspecting a draw.

While a pure strategy Nash equilibrium does not exist due to the possibility of improving overall payoff by adjusting strategies, there may exist a mixed-strategy Nash equilibrium. A mixed-strategy Nash equilibrium occurs when players randomize their actions to achieve the highest expected payoff given their opponent's mixed strategy. In the case of Pecking Order, agents can employ mixed strategies by choosing their actions probabilistically rather than deterministically.

The potential for a mixed-strategy Nash equilibrium in Pecking Order arises from the nature of the game. With imperfect information about the opponent's cards and the dynamic nature of gameplay, players can benefit from randomizing their actions to introduce uncertainty and prevent their opponent from exploiting predictable patterns.

In a mixed-strategy Nash equilibrium, each agent's mixed strategy is best response to the opponent's mixed strategy, ensuring that no player can unilaterally deviate from their strategy to improve their expected payoff. This equilibrium is achieved when the players' mixed strategies are optimal responses to each other.

However, determining the specific mixed-strategy Nash equilibrium in Pecking Order requires detailed analysis and modeling of the game dynamics, probabilities, and strategic interactions. It is an open question and an area of research to explore the existence and properties of mixed-strategy Nash equilibria in Pecking Order.

Therefore, while a pure strategy Nash equilibrium does not exist, the possibility of a mixed-strategy Nash equilibrium suggests that players can benefit from employing probabilistic and adaptive strategies to navigate the complexities of Pecking Order, as explained further in section 2.2.

While theoretically a mixed-strategy approach may be beneficial to playing Pecking Order optimally, in practice, humans often struggle with effectively randomizing their actions (Bacharach and Stahl, 2000; Fehr and Gächter, 2000; Henrich et al., 2001; McKelvey and Palfrey, 1992; Stahl and Wilson, 1995). This difficulty in properly randomizing their choices can lead to sub-optimal outcomes and make it challenging for players to achieve a true mixed-strategy Nash equilibrium, which relies on precise probabilistic calculations and adaptive decision-making. As a result, players may unintentionally exhibit predictable patterns, making them more susceptible to exploitation by their opponents in the game.

2.1.2 Simulation Theory of Mind

The agents in this project utilize a theory of mind approach known as simulation theory of mind. Simulation theory of mind involves predicting an opponent's behavior by simulating what oneself would do in their position (Barlassina and Gordon, 2017). By assuming that their opponent reasons about the game state in the same way they do, an agent expects their opponent to take actions that they themselves would take in that position. Consequently, agents possess beliefs about their opponent's mental states, including beliefs, intentions, and desires, within the context of Pecking Order.

When assuming that their opponent shares the same beliefs, intentions, and desires, an agent concludes that their opponent aims to win the game or, at worst, achieve a draw. At each state of the game, a theory of mind agent (also referred to as a ToM agent) also forms beliefs about their opponent's previous actions and predicts their most probable future actions. A higher-order ToM agent may additionally hold beliefs about what their opponent believes about them. As a result, they can take actions based on what they believe their opponent expects them to have done or do in the future. The level of theory of mind utilization by these agents is further detailed in the following section.

2.2 Agents and their strategies

This study examines strategic decision-making in the card game Pecking Order by analyzing a diverse set of agents with different strategies. The agents include a Random agent, a Levenshtein agent, as well as Zero-Order and First-Order theory of mind agents.

By evaluating the agents' performance against each other, we gain insights into the effectiveness of varying levels of theory of mind within Pecking Order. Subsequent sections will then explore the agents' behaviors, decision-making processes, and outcomes, highlighting the dynamics of strategic interactions in the game.

2.2.1 Random Agent

During each round, the Random Agent selects an available card at random and places it on a randomly selected unoccupied perch. This strategic approach is reminiscent of the one used by rational agents in Limited Bidding, as mentioned in section 2.1.1. By employing random play, the Random Agent aims to mitigate the possibility of their opponents exploiting any predictable patterns in their strategy.

The inclusion of the Random Agent in this study is intended to investigate the potential benefits of employing a random strategy against ToM agents, while also serving as a baseline for comparison to assess the effectiveness of other strategies in the presence of ToM agents.

Note that the Random Agent's strategy is highly exploitable in Pecking Order due to the lack of intention in its actions. Opponents can readily devise counter-strategies and take advantage of the agent's randomness to improve their own outcomes as a result.

2.2.2 Levenshtein Agent

The Levenshtein Agent employs a strategy of generating all possible end-game states, in which both players have played all their cards and thus all cards are face up.

During each turn, this agent compares the current state of the board, including their own perches and the cards occupying them, with each of the generated end-game states. The agent utilizes the Levenshtein similarity metric (Nerbonne et al., 1999), from which its name is derived, to determine the end-game states that exhibit the highest similarity to the current board state. Subsequently, the Levenshtein Agent evaluates the outcomes associated with those selected end-game states, taking into account the winner of each game as well as the final scores of both players. Comparing face-up cards on the board is straightforward, as the value and location of each card can be directly compared to the end-game states. However, for face-down cards, the agent considers the occupancy of the associated perches without directly factoring in the value of their opponents concealed cards. Since the agent knows the value of their own face-down cards, they can factor this into the comparison.

Next, the agent proceeds to filter out any reachable end-game states that do not lead to a victory. In cases where no winning states can be attained from the current board state, the agent instead focuses on states resulting in a tie, as this outcome is preferable to a loss.

The Levenshtein Agent then examines the list of best-matching end-game states, closely resembling the current board state, and simulates placing each of its cards on every unoccupied perch in the current board state. This generates new board states for each valid combination of a card and perch, after which the agent can compare these newly generated states to the each end-game state using the Levenshtein similarity metric. This process enables the agent to determine the action that will yield a board state most closely resembling a winning or tied state for itself. Subsequently, the agent selects an action that ideally leads to a more advantageous outcome given the current state of the board.

In cases where multiple different actions yield the same similarity value and payoff, the agent randomly selects one of these actions. If only one action proves to be advantageous, the agent chooses that action. Finally, if no actions appear to result in a winning or tied state, the agent resorts to playing a random valid action, following a similar approach to the Random Agent. An example of this agents strategy is found in Example 1.

The strategy employed by this agent to assess the value of a particular board state and the value of playing a specific card on a particular perch serves as a foundation for the strategies adopted by the Zero-Order and First-Order ToM agents discussed in sections 2.2.3 and 2.2.4 respectively.

2.2.3 Zero-Order Theory of Mind Agent

The Zero-Order Agent (referred to as the ToM_0 agent from here onwards) evaluates the current board state and his next best actions in a similar manner to the Levenshtein Agent. However, this agent also holds Zero-Order beliefs about the actions of his opponent. These beliefs are represented as a list of tuples, with each entry corresponding to a belief value of a particular card being played on a particular perch in the form: (card, perch, belief value). For instance, the tuple (1, 2, 0.23) represents a 23% belief that the opponent will play card 1 on perch 2. These beliefs are referred to as Zero-Order beliefs.

This agent actually maintains two concurrent belief systems, one for its overall beliefs, and one

Example 1: Consider a Levenshtein Agent playing their final two cards against a Random Agent in a sequential game of Pecking Order. It is the Levenshtein agents turn to play, and the current board state is the following:

Levenshtein Agent	1			4
Random Agent	2	4		

In this board state, the Levenshtein agent has played card 1 on perch 1 and card 4 on perch 4. The Random agent has played card 2 on perch 1 and card 4 on perch 2. Perches 2 and 3 are unoccupied for the Levenshtein agent and perches 3 and 4 are unoccupied for the Random agent. The Levenshtein agent begins by finding the Levenshtein similarity between the current board state and all reachable end-game states. Note that since the Levenshtein agent has no card on perch 2, it does not know that the Random agent has played card 4 on perch 2. Thus, from the perspective of the Levenshtein agent, the board state is the following:

Levenshtein Agent	1			4
Random Agent	2	0		

where 0 represents a face-down card. The greatest Levenshtein similarity value between the current board state (from the Levenshtein agent's perspective) and reachable end-game states is 3, so the Levenshtein agent only considers end-game states with a Levenshtein similarity value equal to 3. The agent then filters out all of the most similar reachable end-game states which do not result in a victory. This leaves the agent with the following two reachable end-game states that result in a victory:

Levenshtein Agent	1	2	3	4
Random Agent	2	4	1	3

Levenshtein Agent	1	3	2	4
Random Agent	2	4	1	3

The Levenshtein agent then simulates playing each of its remaining cards on each of the available perches in the current board state. This generates four new board states reachable from the current board state:

Levenshtein Agent	1	2		4
Random Agent	2	0		

Levenshtein Agent	1		2	4
Random Agent	2	0		

Levenshtein Agent	1	3		4
Random Agent	2	0		

Levenshtein Agent	1		3	4
Random Agent	2	0		

For each of the simulated board states, the Levenshtein agent finds the greatest Levenshtein similarity compared to the two winning end-game states it found earlier. All actions that lead to the same greatest similarity value compared to the end-game states are added to a list. The actions the agent deems to be the most similar based on Levenshtein similarity in this case are playing card 2 on perch 2 and playing card 3 on perch 3. Since there is more than one 'good' action, the agent selects one at random. In this case, the agent opts for playing card 3 on perch 3. The Random agent then plays card 1 on perch 4, leading to the following board state:

Levenshtein Agent	1		3	4
Random Agent	2	4		1

At this point, the Levenshtein and Random agents each only have one possible action left, which determines the final scores of the agents. The final board state results in the Levenshtein agent winning with a total score of 4 and the Random agent losing with a total score of 3:

Levenshtein Agent	1	2	3	4
Random Agent	2	4	3	1

Example 1: Levenshtein Agent Playing Pecking Order Against a Random Agent

for its current beliefs. The overall beliefs persist throughout all games, only being updated at the end of each game when all cards are revealed. These beliefs provide the agent with insight into the general tendencies of its opponent, considering trends across all previous games rather than the current board state. It is important to note that the agent's overall beliefs are randomly initialized upon instantiation and as such are updated over the course of numerous games.

During the game, the agent's current beliefs are updated to better adapt to the ongoing game state. For example, if the opponent plays card 1 on perch 2, it implies that no other cards can be played on perch 2, and card 1 cannot be played elsewhere. Therefore, all current beliefs about other cards on perch 2 are set to zero, and all current beliefs about card 1 being played on other perches are also set to zero.

At the beginning of each game, the agent's current beliefs are set to its overall beliefs and are subsequently updated after each round.

When updating its overall beliefs, this agent incorporates a learning rate specified by the user. The learning rate determines the speed at which the agent adjusts its overall beliefs. In other words, a higher learning rate leads to a greater increase in a belief when encountering a specific card on a certain perch.

Using its Zero-Order beliefs, the agent predicts the values of the opponent's face-down cards and the next most probable card the opponent will play on their next turn based on these beliefs. For instance, if the opponent has a face-down card on perch 2 and this agent strongly believes that the opponent is generally likely to play card 1 on perch 2, it will predict that the opponent's board state includes card 1 on perch 2.

Furthermore, this agent takes an additional step of projecting the opponent's board state one round in the future. For instance, if the opponent currently has no cards on perch 4 and the agent strongly believes that card 3 is likely to be played on perch 4, it will project the opponent's future board state to include card 3 on perch 4. This projection of the future board state enables the agent to proactively choose actions that could potentially result in more favorable outcomes, based on its anticipation of the opponent's likely moves in the next round. By considering these projections, the agent aims to strategically position itself and make informed decisions that align with its beliefs about the opponent's future actions. Note that the predictions of face-down cards and the projected card to be played on the next round result in a single 'projected' board state.

After making predictions and projections, the ToM_0 agent employs the Levenshtein similarity metric to identify comparable winning board states and advantageous actions, similar to the Levenshtein Agent. However, it does so with increased anticipated certainty by substituting face-down cards with the most probable card that the agent believes the opponent will play. Furthermore, projecting the board state one step into the future empowers the agent to take preemptive actions. During each round, when the Levenshtein agent compares its opponent's predicted board state to the final board state, any face-down cards remain hidden, limiting the available information for similarity comparisons and potentially reducing reliability. In contrast, by predicting and projecting the opponent's board state, the ToM_0 agent can establish more precise similarity assessments, enabling it to determine a better action with increased certainty, at least based on its own beliefs.

2.2.4 First-Order Theory of Mind Agent

The First-Order agent (referred to as the ToM_1 agent from here onwards) maintains the same overall and current Zero-Order beliefs as the ToM_0 agent, but it goes further by also incorporating beliefs about its opponent's beliefs, known as First-Order beliefs. Similar to Zero-Order beliefs, this agent maintains both overall and current First-Order beliefs. To form these beliefs, the agent simulates what it would believe if it were in the position of its opponent. Thus overall First-Order beliefs are updated based on the ToM_1 agents own cards placed on their own perches, as this this is what their opponent would use to update their overall beliefs after each game. The ToM_1 agents current First-Order beliefs are updated each turn as if the ToM_1 agents board state was being viewed from the perspective of their opponent. This means that any of the ToM_1 agents cards that would be face-down from the opponents perspective are considered as such, and current First-Order beliefs are adjusted accordingly.

It is important to note that the ToM_1 agent has no direct knowledge of whether its opponent actually possesses any beliefs or instead follows a predetermined strategy. Therefore, this agent maintains a First-Order confidence level c_1, which determines the extent to which its First-Order beliefs influence its decision-making throughout the game. A higher confidence level means the agents First-Order beliefs hold a higher weight, while a lower confidence level means the agents Zero-Order beliefs hold a higher weight.

To determine the next action to take, the ToM_1 agent integrates both its Zero-Order and First-Order beliefs using the First-Order confidence level and a belief integration function, inspired by de Weerd and Verheij (2011) and outlined in section 2.2.5.

When predicting the values of their opponents face-down cards and projecting their opponents

Example 2: Consider a ToM_0 agent playing their final two cards against a Levenshtein Agent in a sequential game of Pecking Order. In this example we consider a scenario where these two agents have already played 25 games against one another, and as such the ToM_0 agent has been able to observe and learn about the tendencies of their opponent. It is the ToM_0 agents turn to play, and the current board state is the following:

ToM_0 agent		3	4	
Levenshtein Agent	1			4

From the perspective of the ToM_0 agent, the board state appears to be the following, where 0s represent face-down cards:

ToM_0 agent		3	4	
Levenshtein Agent	0			0

Over the course of the games played, the Levenshtein agent has had a tendency to play towards the following end-game state:

Levenshtein Agent	1	2	3	4

After each of the previous games played by the agents, the overall Zero-Order beliefs of the ToM_0 agent have been updated to reflect what the agent believes their opponent is most likely to play. At the start of each new game, the ToM_0 agents current Zero-Order beliefs are then set to be equal to the agents overall beliefs. For simplicity, in the game outlined here we will only present the current Zero-Order beliefs of the ToM_0 agent that are greater than 1%:

Current Beliefs				
(Card, Perch, Belief Value)	(1, 1, 0.141)	(2, 2, 0.197)	(3, 3, 0.276)	(4, 4, 0.386)

To select its next action, the ToM_0 agent predicts their opponents face-down cards based on their current beliefs. In this case, the greatest belief values are for cards 1 and 4 being played on perches 1 and 4 respectively, leading to the following board state prediction:

ToM_0 agent		3	4	
Levenshtein Agent	1			4

The ToM_0 agent then projects the next card the Levenshtein agent is most likely going to play. Since the belief value of card 3 being played on perch 3 is greater than the belief value of card 2 being played on perch 2, the ToM_0 agents projection of the current board state one round into the future is:

ToM_0 agent		3	4	
Levenshtein Agent	1		3	4

Using this projected board state, the ToM_0 agent follows the same logic as the Levenshtein agent to select the best action it can play, except it compares the projected board state with reachable end-game states opposed to using the current board state. This allows the ToM_0 agent to more selectively and proactively determine which winning end-game states are reachable from the currently projected board state. The best action determined by the ToM_0 agent is to play card 1 on perch 4. The Levenshtein agent takes the action predicted by the ToM_0 agent, which results in the following board state:

ToM_0 agent		3	4	1
Levenshtein Agent	1		3	4

In the final round both agents play their remaining cards which results in a victory for the ToM_0 agent, after which the ToM_0 agent updates their overall beliefs based on their opponents final board state.

Example 2: ToM_0 Agent Playing Pecking Order Against a Levenshtein Agent future board states, this agent generates both Zero-Order and First-Order projections, based on Zero-Order and First-Order beliefs respectively. This agent also keeps track of all First-Order board state projections during each round throughout the course of a game in order to update its First-Order confidence level. When the opponents face-down cards are revealed, this agent then compares the First-Order projections that they made previously with the actions taken by their opponent and updates their First-Order confidence accordingly.

Similar to the ToM_0 agent, the ToM_1 agent also utilizes a learning rate to regulate the speed at which its beliefs are updated upon encountering specific cards on certain perches.

2.2.5 Belief Adjustment, Belief Integration, Learning Speed and Confidence Level Integration

As previously discussed in sections 2.2.3 and 2.2.4, ToM agents select their actions based on their beliefs and confidence levels. In this section we provide a more detailed explanation of how these agent make use of the learning rate λ to update their beliefs and confidence levels following the outcome of a game.

When a ToM agent is initialized, all of its beliefs are initialized randomly. For a ToM_0 agent, this means all Zero-Order beliefs are randomly initialized, whereas for a ToM_1 agent, both their Zero-Order and First-Order beliefs are randomly initialized. We assume that the ToM_0 agents belief probabilities for each of their opponents actions are non-zero (1) and that the belief probabilities of their opponents actions sum up to 1 (2), taking inspiration from de Weerd et al. (2013):

$$b^{(0)}(c,p) \geq 0 \tag{1}$$

$$\sum_{(c,p)\in(C,P)} b^{(0)}(c,p) = 1 \tag{2}$$

where (c,p) represents a (card, perch) pair, and (C,P) represents the set of all (card, perch) pairs.

After each game is played, a ToM_0 agent updates their overall Zero-Order beliefs by simply incrementing the current belief values for each (card, perch) pair played by their opponent in the game by the learning rate of the agent. After this, the ToM_0 agent recalculates the proportionality of the belief values such that the sum of all belief values equals 1.

For the ToM_1 agent, both its Zero-Order and First-Order beliefs are incremented in the same was as the ToM_0 agent. However it should be noted that the ToM_1 agents First-Order beliefs are updated based on their own cards and perches, as this is what the their opponent would use to update their (assumed) Zero-Order beliefs.

To update its First-Order confidence level c_1, the ToM_1 agent makes use of a confidence level updating function inspired by de Weerd and Verheij (2011):

$$c_1 = \begin{cases} (1-\lambda)\cdot c_1, & \forall p_1 \notin OP \\ \lambda + (1-\lambda)\cdot c_1, & \forall p_1 \in OP \end{cases} \tag{3}$$

where p_1 represents First-Order predictions made by the ToM_1 agent and OP represents the cards on the opponents perches. Following this equation, a ToM_1 agents First-Order confidence level increases with each correct prediction and decreases with each incorrect prediction.

In its decision making process, the ToM_1 agent makes use of integrated beliefs, with are calculated using the belief integration function U, inspired by de Weerd and Verheij (2011). This function combines the ToM_1 agents Zero-Order beliefs b^0 and First-Order beliefs b^1 based on the agents First-Order confidence level c_1.

For a given board state s, the function U determines the integrated beliefs of the agent taking into account the First-Order predicted action $\hat{a}^1(s)$ of its opponent. To form its integrated beliefs, the ToM_1 agent considers each of its current beliefs $(b^0(c,p))$, where (c,p) denotes the (card, perch) pair suspected to be played by the opponent based under the current board state s, after which it weighs these beliefs based on c_1. Formally, the belief integration function is the following:

$$U(b^0, \hat{a}^1(s), c_1)(c,p) =$$

$$\begin{cases} (1-c_1)\cdot b^0(c,p) & \text{if } (c,p) \neq \hat{a}^1(s) \\ c_1 + (1-c_1)\cdot b^0(c,p) & \text{if } (c,p) = \hat{a}^1(s) \end{cases} \tag{4}$$

3 Experimental Setup

To evaluate the effectiveness of the ToM agents presented in section 2.2 across the two game structures, the ToM_0 and ToM_1 agents were paired up with each of the other agents listed across both game structures, after which 400 games per pair were played. After each game, a tally was made denoting the number of wins, losses and draws for each pair. These agent pairs and results can be found in Table 1 in section 4.

The number of games played by each agent pair was determined using the Wald equation (Gudicha et al., 2016) to ensure sufficient games for detecting significant differences in win rates. With an estimated population proportion of 0.5, indicating an expected 50% win rate for competing agents, a sample size of approximately 385 games was calculated. This sample size accounts for a desired level of significance of 0.05 (i.e. 95% confidence level) and a margin of error (E) of 0.05, allowing for

Example 3: Consider a ToM_1 agent playing their final two cards against a ToM_1 Agent in a sequential game of Pecking Order. In this example we consider the same board state as in Example 2 where two agents have already played 25 games against one another, and as such the ToM_1 and ToM_0 agents have both been able to observe and learn about the tendencies of their opponent. In this example we aim to demonstrate how First-Order projections function as this is one of the core differences between the ToM_1 and ToM_0 agents. For the purposes of brevity, we will not outline all beliefs of the ToM_0 agent. It is the ToM_1 agents turn to play, and the board state is the following:

ToM_1 Agent		3	4	
ToM_0 Agent	1			4

From the perspective of the ToM_1 agent, the board state appears to be the following, where 0s represent face-down cards:

ToM_1 Agent		3	4	
ToM_0 Agent	0			0

After each of the previous games, the overall Zero-Order and First-Order beliefs of the ToM_1 agent have been progressively updated, as demonstrated in Example 2. For simplicity, in the game outlined here we present simplified Zero-Order and First-Order beliefs of the ToM_1 agent:

Current Zero-Order Beliefs				
(Card, Perch, Belief Value)	(1, 1, 0.141)	(2, 2, 0.197)	(3, 3, 0.276)	(4, 4, 0.386)

Current First-Order Beliefs				
(Card, Perch, Belief Value)	(2, 1, 0.387)	(3, 2, 0.112)	(4, 3, 0.268)	(1, 4, 0.233)

To select which action to take, the ToM_1 agent begins by simulating their opponents perspective of the board, meaning they predict their own face-down cards using their First-Order beliefs and project their own side of the board one turn into the future based on these beliefs. Note that each First-Order projection made this way is kept track of by the ToM_1 agent to aid in updating their First-Order confidence level later on. Based on the First-Order predictions and projections, the ToM_1 agent then determines what they believe their opponent believes to be the best action they can take. Based on this action, the ToM_1 agent generates a new board state with the suspected action their opponent will take and determines the best action they themself could take, following the same logic as outlined previously for the Levenshtein and ToM_0 agent. In this case, ToM_1 agent suspects the ToM_0 agent to believe that the ToM_1 agent will play card 2 on perch 1 based on First-Order beliefs. The ToM_1 agent then integrates their Zero-Order and First-Order beliefs based on the belief integration function explain in section 2.2.5. For this example, suppose that the ToM_1 agent has a high confidence in their opponent making use of Zero-Order theory of mind. As a result, the ToM_1 agents First-Order beliefs will hold a higher weight in their decision-making process compared to their Zero-Order beliefs. Following their integrated beliefs, the ToM_1 agent determines that the ToM_0 agent may believe that playing card 3 on perch 2 is the best action to take, as this would lead to a lower payoff for the ToM_1 agent opposed to playing card 3 on perch 3. Using this information, the ToM_1 agent could then determine that the best action to take is indeed playing card 2 on perch 1, as regardless of the overall payoff of the game, through controlling perch 1, even if there is a tie, this agent will still win the game. This leads to the following board state:

ToM_1 Agent	2	3	4	
ToM_0 Agent	1	3		4

In the final round both agents play their remaining cards which results in a victory for the ToM_1 agent, after which the ToM_1 agent updates their overall beliefs based on their opponents final board state as well as comparing their First-Order projections with the final outcome of the game state.

Example 3: ToM_1 Agent Playing Pecking Order Against a ToM_0 Agent adequate statistical power in detecting significant differences in win rates.

Both the ToM_0 and ToM_1 agents were assigned a learning rate of 0.4 in order to strike a balance in their belief adjustment speed when attempting to counter their opponents' actions. Setting a lower learning rate may cause the agents' beliefs to adjust too slowly, prolonging the overall learning process and potentially leading to multiple losses and sup-par play before effectively adapting to their opponents' strategies. Conversely, a higher learning rate (close to 1) could result in overreactions to more recent games and lead to unstable behavior and beliefs. By opting for a moderate learning speed of 0.4, the agents could gradually update their knowledge while maintaining the ability to respond effectively to the changing circumstances and new information encountered in each new game.

Upon instantiation, the ToM_1 agents were assigned a confidence level of 0.8, indicating a strong belief that their opponents were utilizing Zero-Order ToM. This initial assumption stems from the absence of direct evidence or prior knowledge regarding their opponent's mental content and reasoning capabilities. By starting with the Zero-Order assumption, ToM_1 agents are able to establish a foundation for reasoning and decision-making while observing their opponent's behaviors. Throughout the course of multiple games, they can adjust their confidence level based on their opponents perceived behaviour. This allows them to determine the weights of their Zero-Order and First-Order beliefs and adjust their strategy based on the emerging patterns and cues displayed by their opponent.

4 Results

The pairing of the ToM_0 and ToM_1 agents with the other listed agents in both game structures resulted in 400 games being played per pair. After each game, the number of wins, losses, and draws for each pair was tallied. The complete set of agent pairs and their respective outcomes can be found in Table 1.

To provide a clearer visual representation of the ToM agents' performance between the two game structures, Figures 2 and 3 present the percentage of games won by the ToM_0 and ToM_1 agents respectively against each of their opponents across the sequential and simultaneous versions of the game.

To assess whether there were significant differences in win rates between the ToM_0 and ToM_1 agents in different game structures, a two-tailed proportion test was conducted. This test compared the win

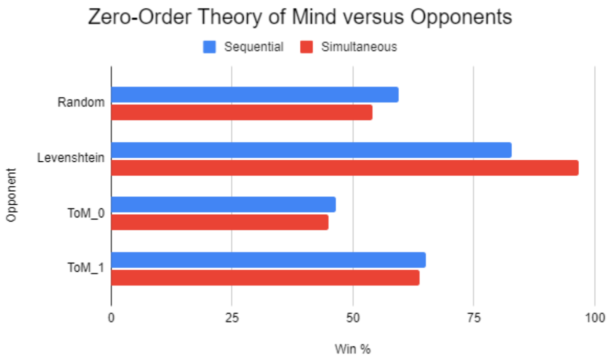

Figure 2: ToM_0 Agent Winning Percentages Versus Opponents Between Game Structures

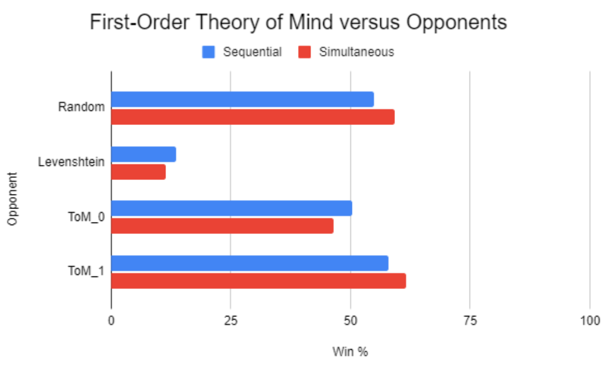

Figure 3: ToM_1 Agent Winning Percentages Versus Opponents Between Game Structures

Game Structure	Agent 1	Agent 2	Agent 1 Win #	Agent 2 Win #	Draw #
Sequential	Zero-Order	Random Action	238	142	20
Sequential	Zero-Order	Levenshtein	331	48	21
Sequential	Zero-Order	Zero-Order	186	199	15
Sequential	Zero-Order	First-Order	260	108	32
Sequential	First-Order	Random Action	220	161	19
Sequential	First-Order	Levenshtein	54	286	60
Sequential	First-Order	Zero-Order	201	183	16
Sequential	First-Order	First-Order	232	154	14
Simultaneous	Zero-Order	Random Action	216	160	24
Simultaneous	Zero-Order	Levenshtein	387	10	3
Simultaneous	Zero-Order	Zero-Order	182	153	65
Simultaneous	First-Order	Random Action	237	147	16
Simultaneous	First-Order	Levenshtein	46	284	70
Simultaneous	First-Order	Zero-Order	186	198	16
Simultaneous	First-Order	First-Order	246	143	11

Table 1: Agent Pair Performance Across Game Structures

rates across the two game structures for each agent's opponents, aiming to detect any notable variations in performance. The null hypothesis (H_0) for this test is that there is no significant difference in win rates of the ToM_0 and ToM_1 agents across the two game structures. The alternative hypothesis (H_a) is that there is a significant difference in win rates of the ToM_0 and ToM_1 agents across the two game structures. The tests were performed at a significance level of $\alpha = 0.05$. The calculated test statistics, including the exact win rates, Z-Scores and p-values for each of the two-tailed proportion tests conducted can be found in Tables 2 and 3.

Multiple chi-squared tests were also conducted to explore the relationship between game structure and win/loss/draw outcomes, considering different combinations of data. The analysis included different combinations of Zero-Order and First-Order win/loss as well as win/loss/draw data. The results of these tests are summarized in Table 4.

The results reveal a significant association between game structure and win/loss/draw outcomes when combining Zero-Order and First-Order data ($\chi^2 = 6.4074$, df = 2, p = 0.04061), indicating the influence of game structure on the overall game outcomes of ToM_0 and ToM_1 agents.

However, analyzing Zero-Order and First-Order win/loss outcomes does not show a significant association with game structure for the combined win/loss performance ($\chi^2 = 1.7226$, df = 1, p = 0.1894). This suggests that game structure may not significantly impact the overall win/loss performance when considering both Zero-Order and First-Order outcomes without the inclusion of draws.

Regarding Zero-Order win/loss/draw outcomes, a significant association with game structure is observed ($\chi^2 = 6.4074$, df = 2, p = 0.04061).

In contrast, no significant association is found between game structure and First-Order win/loss/draw outcomes ($\chi^2 = 1.3114$, df = 2, p = 0.5191), indicating that game structure may not strongly influence the specific win/loss/draw performance in the context of First-Order theory of mind agents.

When examining Zero-Order win/loss outcomes, no significant association with game structure is observed ($\chi^2 = 2.7984$, df = 1, p = 0.09436). Similarly, no significant association is found between game structure and First-Order win/loss outcomes ($\chi^2 = 0.10685$, df = 1, p = 0.7438).

In conclusion, these findings suggest a nuanced relationship between game structure and agent performance, with the impact on win/loss outcomes varying depending on the specific combination of Zero-Order and First-Order data considered as well as whether or not draws are included in the analysis.

5 Discussion

This study aimed to investigate the performance of Zero-Order and First-Order theory of mind agents in sequential and simultaneous game structures of Pecking Order. By comparing win rates among agent pairs and conducting statistical tests, we gained insights into the influence of game structure on the effectiveness of theory of mind reasoning.

Analyzing the data, we conducted two-tailed proportion tests to assess the significance of differences in win rates between the ToM_0 and ToM_1 agents in the two game structures. The results indicated that game structure significantly affected the performance of the ToM_0 agent, with win rates differing significantly between sequential and simultaneous games.

Zero-Order Agent Winning Percentage Across Game Structures					
Opponent		Random	Levenshtein	ToM_0	ToM_1
Game Structure	Sequential	59.5	82.75	46.5	65
	Simultaneous	54	96.75	45	63.75
Z-Score (two-tailed proportion test)		1.57	-6.5278	0.2838	0.3691
p-value (two-tailed proportion test)		0.116	0.00001	0.77948	0.71138

Table 2: Zero-Order Agent Winning Percentage Across Game Structures

First-Order Agent Winning Percentage Across Game Structures					
Opponent		Random	Levenshtein	ToM_0	ToM_1
Game Structure	Sequential	55	13.5	50.25	58
	Simultaneous	59.25	11.5	46.5	61.5
Z-Scores (two-tailed proportion test)		-1.2154	0.8552	1.0612	-1.0093
p-value (two-tailed proportion test)		0.22628	0.38978	0.28914	0.3125

Table 3: First-Order Agent Winning Percentage Across Game Structures

In contrast, the ToM_1 agent's win rates did not exhibit a significant difference between the sequential and simultaneous game structures. Incorporating First-Order beliefs and opponent modeling in the ToM_1 agent's decision-making process appeared to provide a more robust strategy that was less sensitive to changes in game structure.

Furthermore, the findings revealed that turn order may have an impact on the performance of both the Zero-Order and First-Order agents. When the Zero-Order agent played first against the First-Order agent in the sequential variation of Pecking Order, it achieved higher win rates. Similarly, when the First-Order agent played first against the Zero-Order agent in the same context, it attained higher win rates. This observation suggests that turn order may play a role in the effectiveness of theory of mind reasoning in addition to the game structure itself. However, to evaluate whether turn order has a significant impact on the effectiveness of theory of mind reasoning, further research is required.

The significant association between game structure and win/loss/draw outcomes, when combining Zero-Order and First-Order data, suggests an influence of game structure on agent performance. However, when considering only win/loss outcomes, there is evidence of little to no effect of game structure on the effectiveness of theory of mind reasoning.

The findings underscore the importance of incorporating opponent modeling and higher-order theory of mind in decision-making processes, as well as considering the influence of game structure on agent performance. These insights may have implications for the design of intelligent multi-agent systems, enabling the development of more sophisticated and adaptable strategies in dynamic environments.

To further enhance our understanding, future research can explore additional game structures and agent pairings to determine if these may have an impact. Additionally, investigating the combined effects of game structure and turn order on agent behavior and performance can provide insights into optimal strategies for theory of mind reasoning in different game contexts.

Lastly with regards to the initial confidence level and learning rates of the agents, a sensitivity analysis could be conducted to determine the robustness and performance of the agents and how these variations can affect their adaptability, stability, and overall effectiveness in responding to their opponents' strategies. By systematically evaluating different learning rates and confidence levels, the analysis can provide valuable insights into the optimal parameters that maximize the agents' ability to adjust their strategies and achieve superior gameplay outcomes.

6 Conclusions

This study analyzes the performance of Zero-Order and First-Order theory of mind agents in both sequential and simultaneous game structures of Pecking Order. By conducting win rate analysis and statistical tests, we investigated how game structure affects the performance of agents utilizing theory of mind.

The results reveal slight variations in agent performance across game structures, highlighting the effectiveness of different strategies in the context of Pecking Order. The Zero-Order agent's performance was significantly influenced by game structure, while the First-Order agent appeared to be less influenced overall. Turn order also seemed to a play role, with

Agent Data	Chi-Squared (X^2)	Degrees of Freedom (df)	p-value
Zero-Order + First-Order Win/Loss/Draw	6.4074	2	0.04061
Zero-Order + First-Order Win/Loss	1.7226	1	0.1894
Zero-Order Win/Loss/Draw	6.4074	2	0.04061
First-Order Win/Loss/Draw	1.3114	2	0.5191
Zero-Order Win/Loss	2.7984	1	0.09436
First-Order Win/Loss	0.10685	1	0.7438

Table 4: Chi-Squared Analysis Across Game Structures

higher win rates observed when the Zero-Order agent played first against the First-Order agent and vice versa, however this requires further investigation to gain more conclusive results.

This study aimed to make meaningful contributions to the field of artificial intelligence and multi-agent systems by investigating the intricate interplay between game structure and the utilization of theory of mind. By gaining a deeper understanding of this relationship, we can pave the way for the development of more sophisticated implementations of theory of mind reasoning in various contexts. These findings have the potential to extend beyond the realm of games, offering insights and applications that can benefit the broader field of artificial intelligence and multi-agent systems as a whole.

References

Bacharach, M. and Stahl, D. O. (2000). Variable-frame level-n theory. *Games and Economic Behavior*, 32(2):220–246.

Barlassina, L. and Gordon, R. M. (2017). Folk psychology as mental simulation. In Zalta, E., editor, *The Stanford Encyclopedia of Philosophy*. Center for the Study of Language and Information, Stanford.

de Weerd, H., Verbrugge, R., and Verheij, B. (2013). Higher-order social cognition in the game of rock-paper-scissors: A simulation study. In *AAMAS'12: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pages 1195–1196.

de Weerd, H. and Verheij, B. (2011). The advantage of higher-order theory of mind in the game of limited bidding. In *Proceedings Workshop 'Reasoning about other Minds', CEUR Workshop Proceedings*, volume 751, pages 149–164.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Garfield, R. (1998). Pecking order. [Board game].

Gudicha, D. W., Tekle, F. B., and Vermunt, J. K. (2016). Power and sample size computation for wald tests in latent class models. *Journal of Classification*, 33:30–51.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78.

McKelvey, R. D. and Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica: Journal of the Econometric Society*, pages 803–836.

Nerbonne, J., Heeringa, W., and Kleiweg, P. (1999). Edit distance and dialect proximity. *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, 15.

Stahl, D. O. and Wilson, P. W. (1995). On players models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.