Towards Transparent Computational Models of Theory of Mind in Collaborative Environments

Carmine Grimaldi[1,*], Silvia Rossi[1,*]

[1]University of Napoli Federico II, Napoli, Italy

Abstract

In complex multi-agent environments, constructing insightful models of counterparts is paramount for intelligent systems. Central to this challenge is the adoption of a transparent Theory of Mind (ToM), encompassing unobservable mental states like desires, beliefs, and intentions. This paper investigates the role of ToM in collaborative contexts, particularly emphasizing its application in the Hanabi card game. By leveraging transparent decision-making interactions, we aim to investigate how ToM shapes agents' efficacy in coordinating within collaborative frameworks. Our approach prioritizes transparency by comparing logic-based and decision-tree methodologies in modelling ToM reasoning. We propose utilizing the Hand Card Information Completion module to generate beliefs about players' hands, integrated into both approaches. In the logic-based method, reasoning is conducted through logical inference to infer optimal actions based on game history and contextual cues. Conversely, the decision trees approach decomposes decision-making into hierarchical levels, facilitating efficient navigation of the decision space. Our goal is to evaluate the effectiveness and comprehensibility of these approaches, offering a deeper understanding of their strengths and weaknesses in fostering transparent and proficient interactions in collaborative environments.

Keywords

Theory of Mind, Collaborative Environments, Transparent Decision-Making, Computational Models, Hanabi Card Game

1. Introduction

In collaborative environments, understanding the minds of others is paramount for the success of shared interactions and activities. Theory of Mind, referring to the ability to attribute mental states such as beliefs, desires, and intentions to others, plays a crucial role in this context. Figure 1 provides a visual representation of how individuals form mental models of others' mental states. The illustration depicts two individuals, each considering the other's thoughts and beliefs, creating a nested structure of mental states. In collaborative environments, this ability is crucial as it allows agents to predict and interpret the actions and intentions of their collaborators, leading to more effective and synchronized teamwork. However, applying ToM in complex environments, such as the Hanabi card game, presents unique challenges. Hanabi requires players to infer their teammates' hidden information while strategically managing their own actions. To address these challenges, computational models for ToM offer valuable insights into agents' decision-making processes. In this paper, we delve into the application of ToM in Hanabi, focusing on two distinct computational approaches—logic-based and decision trees. Our aim is to evaluate and compare the effectiveness and transparency of these models in capturing ToM reasoning. Additionally, we emphasize the importance of transparency in decision-making processes within collaborative settings, aiming to enhance the interpretability and understanding of our models' outputs. To provide a comprehensive understanding of ToM in collaborative environments, we briefly introduce state-of-the-art computational models.

WOA 2024: 25th Workshop "From Objects to Agents", July 8-10, 2024, Forte di Bard (AO), Italy

*Corresponding author.
carmine.grimaldi@studenti.unina.it (C. Grimaldi); silvia.rossi@unina.it (S. Rossi)

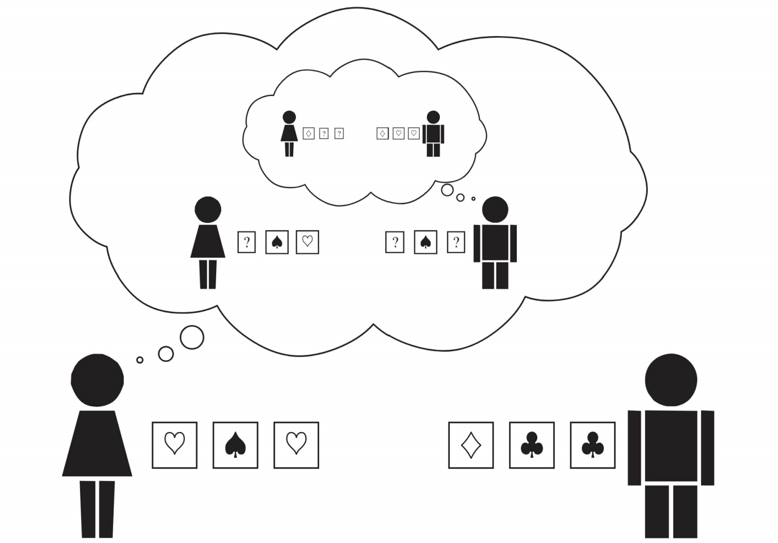

Figure 1: Theory of mind graphical representation: having a mental model of other agents' mental models.

2. Transparency and Theory of Mind Models

The exploration of explanation has captivated philosophers for ages, aiming to discern its essence and the intricacies of its structure and function. More recently, psychologists delved into how humans rationalize others' behaviours and the general process of explanation generation and evaluation [1, 2, 3, 4]. Within AI, explanation has also been researched extensively with early work including a variety of logic-based and probabilistic approaches to abductive inference or so-called inference to the best explanation including early works like [5, 6, 7]. The resurgence of interest in explanation within AI, often framed as Explainable AI, stems from the necessity of providing interpretable justifications for decision-making in an opaque machine and deep learning systems [8, 9, 10]. The ability to provide descriptions and explanations of other's behaviour intersects with the Theory of Mind, a fundamental aspect in understanding how individuals perceive and analyze each other's reasoning processes. Premack and Woodruff explained this term as referring to an understanding of others and oneself as having mental states [11]. Crucially, ToM becomes even more vital in partially observable environments, where agents must reason about each other's beliefs and intentions based on limited observations. Recent studies have suggested that machine ToM may also emerge in Large Language Models (LLMs) like GPT-4 [12, 13]. However, more systematic evaluations have indicated that the apparent ToM capacities in LLMs are not yet as robust as those in humans [14, 15, 16], often failing to pass trivial variations of common tests [17].

In the literature, there have been two primary approaches to engineering machine ToM: end-to-end methods like Theory of Mind neural networks [18], and model-based methods such as Bayesian inverse planning [19, 20]. End-to-end methods aim to learn ToM capabilities directly from data without explicitly modelling mental states or reasoning processes. These approaches leverage deep learning architectures to extract patterns and relationships from raw data, enabling agents to infer the intentions and mental states of others implicitly. While these methods offer the advantage of flexibility and scalability, they may lack interpretability compared to more explicitly structured approaches. In contrast, the goal of Bayesian approaches to engineering ToM lies in probabilistically modelling the mental states and reasoning processes of other agents. By incorporating Bayesian inference, these methods aim to infer the beliefs, desires, and intentions of agents based on observed behaviour and environmental cues. This approach allows for a principled framework to reason about uncertainty and make decisions in interactive settings, particularly in contexts with limited observability or ambiguous information. However, Bayesian approaches also face several limitations. One significant limitation is the computational complexity associated with Bayesian inference, especially in scenarios with large state or action spaces. Inference in Bayesian models often requires iterative updating of beliefs based on new evidence, which can be computationally demanding and prohibitive in real-time or resource-constrained environments. Additionally, Bayesian approaches rely on accurate specification of prior beliefs and likelihood functions, which may be challenging to obtain in practice.

Both end-to-end and Bayesian approaches primarily focus on unimodal data and straightforward domains. Baker and Tanenbaum [19, 21] present frameworks for representing and addressing challenges posed by environments with limited observability. They demonstrate that the Bayesian Theory of Mind harmonizes effectively with how individuals perceive and analyze each other's reasoning processes.

Meanwhile, strategies showcased by Hanabi-playing agents in [22], illustrate the effectiveness of integrating Agent Modelling with Monte Carlo tree search (MCTS) [23]. The benefits of the MCTS approach lie in its ability to handle complex decision-making problems by balancing exploration and exploitation effectively. By performing random sampling through simulations and storing action statistics, MCTS can make informed decisions in each iteration, leading to improved performance over time [23]. This approach has proven particularly successful in combinatorial games, where it has become a state-of-the-art technique. In [24, 25], researchers utilize the MCTS algorithm to deduce the opponent's model. However, MCTS also has its limitations. In more complex games with high branching factors or real-time constraints, its efficiency may be compromised. Furthermore, for some games such as Go, a single combinatorial approach does not always lead to a reliable evaluation of the game states [26].

ToM capabilities and their impact on the effectiveness of interaction have been explored in [27, 28]. In these works, ToM is harnessed to minimize communication, with observations indicating significant benefits of the ToM-based approach over alternative baseline algorithms. However, [28] operates under the assumption of complete observability. In [29], the authors propose and evaluate the performance of a Level 1 Theory of Mind agent, which utilizes agent modelling to construct a cognitive representation of another agent subject to constraints on communication and observability. This agent not only considers observed information and received communication from its counterpart but also accounts for missing environmental information, facilitating an understanding of the other agent's belief state. Their contributions include the utilization of the Monte-Carlo Tree Search algorithm for constructing belief trees to facilitate selective communication among agents, along with the presentation of mathematical formulations for updating beliefs following communication actions.

Logic-based approaches, such as those leveraging Dynamic Epistemic Logic (DEL) [30, 31], focus on formalizing the knowledge and belief dynamics within strategic interactions. These methods aim to capture the process of adjusting one's beliefs in response to new information and information exchange among agents in a structured and logical manner. Using formal logic, such as DEL, provides a rigorous framework for representing and reasoning about agents' beliefs, knowledge, and actions. To formalize belief manipulation in strategic interactions, [32] contributes significantly by providing a practical approach to specifying belief manipulation actions in games. They focus on situations where agents' knowledge is affected asymmetrically, a common scenario in various domains ranging from card games like poker to secure information transmission. By leveraging DEL as a theoretical foundation, the authors introduce Ostari, which provides an implementation of a particular flavour of DEL, presented by Baltag [33]. Ostari simplifies the expression of belief manipulation actions, thus facilitating their practical utilization for modelling complex scenarios, such as popular card games. Notably, the work addresses the inherent complexity of DEL by offering concise representations of actions while retaining its expressive power.

One major limitation of logic-based approaches is the computational complexity involved in reasoning with formal logic, especially in scenarios with large or complex state spaces. Dynamic Epistemic Logic and similar formalisms often require significant computational resources to perform belief updates and reason about agents' knowledge and beliefs accurately. Furthermore, these approaches may face difficulties in handling incomplete or uncertain information effectively. In many real-world scenarios, agents have limited or imperfect information about the environment and other agents' mental states. Representing and reasoning about uncertainty in a logical framework can be challenging and may

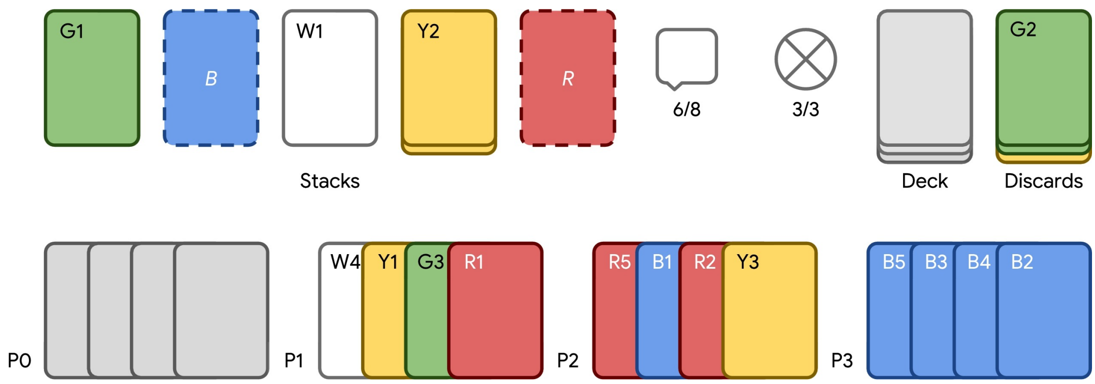

Figure 2: Example of a four-player Hanabi game from the point of view of player 0. Player 1 acts after player 0 and so on.

require additional mechanisms, such as probabilistic extensions to logic or hybrid approaches combining logic with other formalisms.

Overall, while significant progress has been made in understanding and engineering explanation and Theory of Mind in artificial systems, there remains ample room for further exploration and refinement, particularly in addressing the complexities of real-world environments and achieving human-comparable performance.

2.1. The Hanabi Card Game

Hanabi is a cooperative card game designed for two to five players. Each player is dealt a hand of four cards, or five when playing with two or three players. These cards display both a rank, ranging from 1 to 5, and a colour, including red, green, blue, yellow, and white. The game deck comprises a total of 50 cards, with 10 cards of each color: three of rank 1, two of rank 2, 3, and 4, and a single card of rank 5. The objective of Hanabi is to strategically play cards to construct five stacks, each representing a different colour, starting from rank 1 and concluding at rank 5. What distinguishes Hanabi is its innovative gameplay mechanic: players can only view the hands of their fellow players, not their own. In this cooperative setting, players must rely on communication and deduction to coordinate their moves effectively. In Hanabi, the challenge lies in players' limited ability to convey information to one another and the finite amount of information that can be exchanged throughout the game. This dynamic creates a tension-filled experience where players must carefully manage their resources and communicate efficiently to achieve success.

At the beginning of the game, players have access to two types of tokens:

- **Information Tokens:** These tokens are used by players to give clues to their teammates about the cards they hold. Players can spend an information token to either give a clue about a specific colour or a specific number to a teammate. Each game starts with 8 available information tokens.
- **Fuse Tokens:** Fuse tokens represent the fuse of the fireworks. A fuse token is discarded when a player makes a mistake, such as playing a card out of order or giving incorrect information. If all three fuse tokens are discarded, the game ends immediately, and players lose.

Play proceeds around the table; each turn, a player must take one of the following actions:

- **Give information:** During their turn, the player in action has the option to give a hint to any other participant. This hint involves selecting either a specific rank or colour and then pointing out to the chosen player all the cards in their possession that match the designated rank or colour. However, it's important to note that only ranks and colours present in the player's hand can be indicated. For instance, in Figure 2, player 0 (the active player) might tell player 2, without revealing the exact card, something like, "Your first and third cards are red," or "Your first card is

a 5." To maintain the challenge of the game, hints are limited, and they consume an information token each time a hint is given. Once all information tokens are depleted, no further hints can be provided, and the player must proceed with either playing a card or discarding one.
- **Discard a card:** The player chooses a card from their hand and adds it to the discard pile, then draws a card to replace it. The discarded card is out of the game and can no longer be played. Discarding a card restores one information token.
- **Play a card:** The player takes a card from their hand and places it face up in front of them. For example, in Figure 2, player 1's action would be successful if they played their red 1, forming the beginning of the red stack. Considering that there can only be one firework of each colour, that cards must be played on fireworks in ascending numerical order (1, 2, 3, 4, 5) and that in each firework there can only be one card of each value (5 cards in total), there are then two possibilities:
  - if the card can start, or be added to a firework, it is placed face-up on that firework's stack;
  - if the card cannot be added to a firework, it is discarded and a fuse token is placed.

  In either case, the player then draws a new card, without looking at it, and adds it to their hand.

Hanabi is typically played as a single-team game, where all players collaborate to achieve the game's objectives. However, players may choose to divide themselves into smaller teams within the group, as long as they follow the cooperative spirit of the game. The game can conclude in one of three ways:

1. If the group successfully plays cards to complete all five stacks, they achieve victory together.
2. If three lives have been lost due to fuse tokens being discarded, the game ends immediately, and all players lose.
3. If a player draws the last card from the deck, and every player has taken their final turn, the game ends. In this case, the group earns points based on the cards in the completed stacks.

If the game concludes before three lives are lost, the group earns one point for each card in every stack, with a maximum possible score of 25. However, if the game ends due to the loss of three lives, the score is set at 0.

The Hanabi card game is a meaningful benchmark because it requires Theory of Mind reasoning and challenges an agent's decision-making ability in a partially observable and cooperative setting. The authors in [34] proposed two innovative plug-in modules that can be applied to general reinforcement learning agents. The Hand Card Information Completion module is designed to model other agents' mental states and complement environment information. Meanwhile, the Goal-Oriented Intrinsic Reward module encourages agents' exploration and collaboration. In the domain of reinforcement learning methodologies, where achieving explainability remains a formidable challenge, significant approaches have surfaced. In [35], the Authors present Policy Belief Iteration (Pb-It), a method aimed at learning implicit communication protocols within the card game bridge. Here, agents must convey information about their hands to partners through bidding actions. Pb-It introduces a belief update function that forecasts the partner's private information based on their actions. Moreover, an auxiliary reward incentivizes agents to communicate bids that maximize uncertainty reduction for their partner. However, Pb-It's reliance on access to private information during training limits its practicality. In [36], the Authors introduce the Actor-Critic-Hanabi-Agent (ACHA), employing deep neural networks for parametrization. ACHA leverages the Importance Weighted Actor-Learner to circumvent stagnant gradients and employs population-based training for hyperparameter optimization. In [37], the Authors present Rainbow, a fusion of cutting-edge Deep Q-Network (DQN) techniques, including Double DQN, Noisy Networks, Prioritized Experience Replay, and Distributional Reinforcement Learning. However, these techniques lack transparency as they do not allow for the generation of explanations. Finally, in [38] the Authors proposed the Bayesian Action Decoder (BAD), a multi-agent learning method that uses an approximate Bayesian update to obtain a public belief that conditions on the actions taken by all agents in the environment. The BAD offers a comprehensive explanation of its framework, detailing components like public belief, deterministic partial policies characterized by deep neural networks, and the approximate Bayesian update process. Yet, achieving complete transparency entails being able to track and comprehend each specific step and calculation the model undertakes to decide at any given moment. While the BAD presents a coherent conceptual framework for decision-making, full transparency may be elusive due to its reliance on deep neural networks, which are trained to approximate and parameterize deterministic partial policies, allowing agents to map private observations to environmental actions. In essence, using neural networks adds complexity to understanding the model's inner workings.

3. A Proposed Approach

In [34] the authors introduce a groundbreaking approach to Hanabi using Deep Q-Networks in Reinforcement Learning, presenting a novel Hand Card Information Completion (HCIC) module that yields invaluable insights. The HCIC module operates by synthesizing global action history and observable card data, emulating human-like inference processes. Through a Transformer-based architecture, it extrapolates the agent's own hand information, bridging the gap between partial observations and informed decision-making. The experiments, conducted within the Rainbow DQN framework, demonstrate notable enhancements in performance, particularly evident in games involving three or four players. However, the utilization of Deep Q-Networks introduces a layer of complexity that hinders explainability, thus obscuring the rationale behind the agent's decisions. We aim to start from the HCIC module and utilize it to generate beliefs about the cards held by each player. This approach provides insight into what each player perceives their hand to be. Subsequently, we will introduce two different Theory of Mind approaches, leveraging the information extracted from the HCIC module. We will then compare the results with existing literature and those obtained by [34] using the Rainbow DQN framework. The two approaches we propose are based on logic-based and decision-tree methodologies, which offer the advantage of being transparent and easily interpretable.

3.1. Logic-based approach

In the logic-based approach, we utilize ProbLog, a probabilistic logic programming language that extends Prolog with probabilities [39, 40, 41]. Through ProbLog, we can define a Knowledge Base to model Hanabi, where possible actions such as playing a card or giving a hint are represented. Additionally, probabilities of possessing a certain card for a given player are introduced. These probabilities are initialized and continuously updated using the HCIC module. Subsequently, the idea is to define inference rules for selecting the best action to take, given the game's history.

To provide a clearer explanation, our approach involves extracting metarules from the HCIC module to enhance inference in ProbLog. For instance, we assess the likelihood of suggesting a colour as a favourable action considering factors such as uncertainty and other relevant considerations. In the context of Hanabi, a player is considered uncertain about a colour if they do not have sufficient knowledge to confirm or deny the presence of cards of that colour in their hand. This is expressed as:

$$\begin{aligned}&\text{check\_color\_uncertainties(Player, [Index | RemainingIndexes])} \equiv \\ &\quad \Big(\text{Index} \geq 1 \land \text{Index} \leq 5 \land \\ &\qquad \Big(\text{color\_uncertainty(Player, Index)} \lor \text{check\_color\_uncertainties(Player, RemainingIndexes)}\Big)\Big)\end{aligned}$$

$$\begin{aligned}\text{color\_uncertainty}(Player, Index) \equiv \Big(&\text{at}(Player, Index, Card) \land \\ &\big((\text{color}(Card, \text{red}) \land \neg\text{red\_idx}(Player, Index)) \\ &\quad \lor (\text{color}(Card, \text{blue}) \land \neg\text{blue\_idx}(Player, Index))\big)\Big)\end{aligned}$$

The `red_idx(Player, Index)` and `blue_idx(Player, Index)` predicates represent the probability that the player believes to have a red or blue card, respectively, at a specific index of their hand.

Since we are interested in the player's uncertainty, we use the complement of these probabilities. The `color_uncertainty` formula states that a player is uncertain about the colour of a card at a given index if:

1. The card exists at that index (`at(Player, Index, Card)`).
2. For a red card, the player is uncertain if `colour(Card, red)` is true but they do not know it (`¬red_idx(Player, Index)`).
3. For a blue card, the player is uncertain if `colour(Card, blue)` is true but they do not know it (`¬blue_idx(Player, Index)`).

Combining these conditions with a logical OR ($\vee$), the formula effectively captures the scenarios where the player lacks definite knowledge about the colour of the card at a specific index, thereby indicating their uncertainty.

1. The card exists at that index (`at(Player, Index, Card)`).
2. For a red card, the player is uncertain if `colour(Card, red)` is true but they do not know it (`¬red_idx(Player, Index)`).
3. For a blue card, the player is uncertain if `colour(Card, blue)` is true but they do not know it (`¬blue_idx(Player, Index)`).

Combining these conditions with a logical OR ($\vee$), the formula effectively captures the scenarios where the player lacks definite knowledge about the colour of the card at a specific index, thereby indicating their uncertainty.

In addition, we integrate further contextual cues to refine probabilities, including the number of available information tokens, the count of discarded cards, and other relevant factors. This approach ensures that our inference process dynamically adapts to the current state of the game, allowing for more accurate adjustments in probabilities.

Continuing the example of suggesting a colour, in that case, we adjust the probability of needing to take such action based on the following:

- The number of playable cards of the hinted colour (*NumColoredCards*).
- The number of cards remaining in the deck (*DeckSize*).
- The number of available hint tokens (*Hints*).

These adjustments are represented by probabilistic facts that assume a probability value depending on the argument they receive:

$$\begin{aligned}
&\text{alter_hint_color_deckSize}(DeckSize)\\
&\text{alter_hint_color_playableCards}(NumColoredCards)\\
&\text{alter_hint_color_hintTokens}(Hints)
\end{aligned}$$

The final probability of suggesting a hint of colour is computed by combining the uncertainties and contextual adjustments. This involves:

$$\begin{aligned}
&\text{hint_color_probability}(Player, HintedPlayer, Color) \equiv\\
&\Big(\text{number_of_hinted_playable_colored_cards}(HintedPlayer, Color, NumColoredCards)\wedge\\
&\quad \text{remaining_deck_size}(DeckSize) \wedge \text{hints_available}(Hints)\wedge\\
&\quad \text{findall}(Index, (\text{at}(OtherPlayer, Index, Card),\\
&\qquad \text{color}(Card, CardColor), CardColor = Color), Indices)\wedge\\
&\quad \Big(\text{check_color_uncertainties}(HintedPlayer, Indices)\\
&\quad \vee \text{alter_hint_color_playableCards}(NumColoredCards)\\
&\quad \vee \text{alter_hint_color_deckSize}(DeckSize)\\
&\quad \vee \text{alter_hint_color_hintTokens}(Hints)\Big)\Big)
\end{aligned}$$

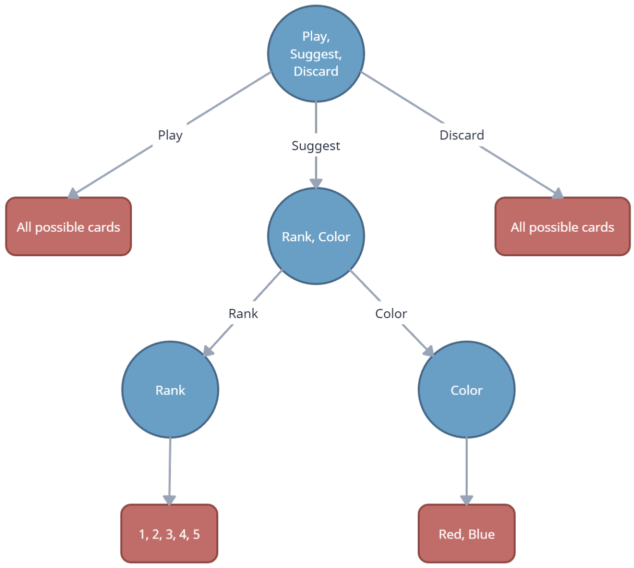

Figure 3: Decision trees diagram for a simplified Hanabi game with only two colours. There are a total of six decision trees: 1. Play, Suggest, Discard; 2. Play; 3. Discard; 4. Rank, Colour; 5. Rank; 6. Colour.

3.2. Decision trees approach

Another approach we propose relies on decision trees. Initially, we create a dataset by combining the outputs of the HCIC module for all players with additional contextual information. This dataset is generated within the Hanabi Learning Environment [42], where agents engage in gameplay with each other, providing multiple scenarios for training purposes. The dataset comprises various crucial pieces of information, such as the current state of the game, the cards held by each player, the available hints, and the observed actions.

Given computational constraints, employing a single decision tree to determine the optimal action would be excessively burdensome. Consequently, we opt to decompose this task into subtasks, enabling sequential decision-making.

Within this framework, we construct multiple decision trees operating at different hierarchical levels (see Figure 3). For instance, a higher-level decision tree receives inputs from the HCIC module, presenting the optimal choice among playing a card, discarding a card, or providing a hint. Subsequently, based on the decision given by this primary decision tree, another decision tree corresponding to the selected action is activated. If, for example, the optimal action is determined to be playing a card, then the decision tree responsible for selecting which is the best card to play is invoked. Each decision tree is trained on specific subsets of the dataset tailored to its task. For instance, there's a decision tree trained solely on determining which card to play, another focused on discerning whether to give a hint of rank or colour and yet another specialized in suggesting which colour to hint. This specialization ensures that each decision tree is trained with relevant input features and output labels pertinent to its specific decision-making task.

By employing this hierarchical structure, we mitigate computational complexity while ensuring effective decision-making. Each decision tree operates within a specific domain of decision space, enabling efficient navigation and decision optimization within the Hanabi game environment. Additionally, this hierarchical framework allows for easier interpretability and transparency, as decision processes are delineated into distinct levels of abstraction, thereby facilitating analysis and evaluation.

4. Conclusions

In this article, our focus is on enhancing the discourse surrounding computational models for the Theory of Mind by fostering an approach which emphasizes transparency and explicability. Our primary aim is to evaluate various methodologies, particularly comparing logic-based and decision trees approaches, to gauge their effectiveness and comprehensibility. By prioritizing transparency, we seek to enhance the understandability and interpretability of our models, thus aligning them more closely with human cognitive processes.

We will delve into a detailed comparison of the two models to discern their respective strengths and weaknesses. To achieve this, we will employ various metrics and indicators of transparency. These may include but are not limited to, the clarity and comprehensiveness of model documentation, the degree of insight provided into the decision-making process, and the accessibility of underlying assumptions and parameters. Furthermore, we will assess the models' robustness to input variations and their ability to accommodate uncertainty and ambiguity inherent in real-world scenarios. Through these evaluations, we aim to provide an understanding of the trade-offs between the logic-based and decision trees approaches, highlighting their applicability across various contexts and scenarios.

ACKNOWLEDGMENT

This study received funding from the European Union - Next-GenerationEU - National Recovery and Resilience Plan (NRRP) – MISSION 4 COMPONENT 2, INVESTIMENT N. 1.1, CALL PRIN 2022 PNRR D.D. 1409 14-09-2022 – ADVISOR CUP N.E53D23016260001.

References

[1] D. Hilton, Conversational processes and causal explanation, Psychological Bulletin 107 (1990) 65–81. doi:10.1037/0033-2909.107.1.65.

[2] C. Antaki, I. Leudar, Explaining in conversation: towards an argument model., European Journal of Social Psychology 22 (1992) 181–194. doi:10.1002/ejsp.2420220206.

[3] B. R. Slugoski, M. Lalljee, R. Lamb, G. P. Ginsburg, Attribution in conversational context: Effect of mutual knowledge on explanation-giving, European Journal of Social Psychology 23 (1993) 219–238. URL: https://api.semanticscholar.org/CorpusID:145101157.

[4] B. Malle, J. Knobe, M. O'Laughlin, G. Pearce, S. Nelson, Conceptual structure and social functions of behavior explanations: Beyond person-situation attributions, Journal of personality and social psychology 79 (2000) 309–26. doi:10.1037//0022-3514.79.3.309.

[5] H. E. Pople, On the mechanization of abductive logic, in: Proceedings of the 3rd International Joint Conference on Artificial Intelligence, IJCAI'73, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1973, p. 147–152.

[6] E. Charniak, D. Mcdermott, Introduction to Artificial Intelligence, Addison-Wesley, 1986.

[7] H. J. Levesque, A knowledge-level account of abduction, in: International Joint Conference on Artificial Intelligence, IJCAI'89, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989, p. 1061–1067. URL: https://api.semanticscholar.org/CorpusID:261943411.

[8] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, 2017. arXiv:1702.08608.

[9] W. Samek, T. Wiegand, K.-R. Müller, Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, 2017. arXiv:1708.08296.

[10] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G.-Z. Yang, Xai—explainable artificial intelligence, Science Robotics 4 (2019) eaay7120. doi:10.1126/scirobotics.aay7120.

[11] D. Premack, G. Woodruff, Does a chimpanzee have a theory of mind, Behavioral and Brain Sciences 1 (1978) 515 – 526. doi:10.1017/S0140525X00076512.

[12] M. Kosinski, Evaluating large language models in theory of mind tasks, 2024. arXiv:2302.02083.

[13] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, Y. Zhang, Sparks of artificial general intelligence: Early experiments with gpt-4, 2023. arXiv:2303.12712.
[14] M. Sap, R. LeBras, D. Fried, Y. Choi, Neural theory-of-mind? on the limits of social intelligence in large lms, 2023. arXiv:2210.13312.
[15] N. Shapira, M. Levy, S. H. Alavi, X. Zhou, Y. Choi, Y. Goldberg, M. Sap, V. Shwartz, Clever hans or neural theory of mind? stress testing social reasoning in large language models, 2023. arXiv:2305.14763.
[16] M. Sclar, S. Kumar, P. West, A. Suhr, Y. Choi, Y. Tsvetkov, Minding language models' (lack of) theory of mind: A plug-and-play multi-character belief tracker, 2023. arXiv:2306.00924.
[17] T. Ullman, Large language models fail on trivial alterations to theory-of-mind tasks, 2023. arXiv:2302.08399.
[18] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. M. A. Eslami, M. Botvinick, Machine theory of mind, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 of *Proceedings of Machine Learning Research*, PMLR, 2018, pp. 4218–4227. URL: https://proceedings.mlr.press/v80/rabinowitz18a.html.
[19] C. Baker, J. Jara-Ettinger, R. Saxe, J. Tenenbaum, Rational quantitative attribution of beliefs, desires and percepts in human mentalizing, Nature Human Behaviour 1 (2017) 0064. doi:10.1038/s41562-017-0064.
[20] T. Shu, A. Bhandwaldar, C. Gan, K. A. Smith, S. Liu, D. Gutfreund, E. Spelke, J. B. Tenenbaum, T. D. Ullman, Agent: A benchmark for core psychological reasoning, 2021. arXiv:2102.12321.
[21] C. L. Baker, R. Saxe, J. B. Tenenbaum, Action understanding as inverse planning, Cognition 113 (2009) 329–349. doi:10.1016/j.cognition.2009.07.005.
[22] J. Walton-Rivers, P. R. Williams, R. Bartle, D. Perez-Liebana, S. M. Lucas, Evaluating and modelling hanabi-playing agents, in: 2017 IEEE Congress on Evolutionary Computation (CEC), IEEE Press, 2017, p. 1382–1389. URL: https://doi.org/10.1109/CEC.2017.7969465. doi:10.1109/CEC.2017.7969465.
[23] M. Świechowski, K. Godlewski, B. Sawicki, J. Mańdziuk, Monte carlo tree search: a review of recent modifications and applications, Artif. Intell. Rev. 56 (2022) 2497–2562. URL: https://doi.org/10.1007/s10462-022-10228-y. doi:10.1007/s10462-022-10228-y.
[24] J. Goodman, S. Lucas, Does it matter how well i know what you're thinking? opponent modelling in an rts game, in: 2020 IEEE Congress on Evolutionary Computation (CEC), IEEE Press, 2020, p. 1–8. URL: https://doi.org/10.1109/CEC48606.2020.9185512. doi:10.1109/CEC48606.2020.9185512.
[25] S. Barrett, P. Stone, S. Kraus, Empirical evaluation of ad hoc teamwork in the pursuit domain, in: The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2, AAMAS '11, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2011, p. 567–574.
[26] A. Fabbri, F. Armetta, E. Duchêne, S. Hassas, Knowledge complement for monte carlo tree search: An application to combinatorial games, in: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, 2014, pp. 997–1003. doi:10.1109/ICTAI.2014.151.
[27] Y. Wang, F. Zhong, J. Xu, Y. Wang, Tom2c: Target-oriented multi-agent communication and cooperation with theory of mind, 2022. arXiv:2111.09189.
[28] M. C. Buehler, T. H. Weisswange, Theory of mind based communication for human agent cooperation, in: 2020 IEEE International Conference on Human-Machine Systems (ICHMS), 2020, pp. 1–6. doi:10.1109/ICHMS49158.2020.9209472.
[29] P. Desai, R. Singh, T. Miller, Theory of mind for selective communication and enhanced situational awareness, in: AIAC 2023: 20th Australian International Aerospace Congress: 20th Australian International Aerospace Congress, Engineers Australia Melbourne, 2023, pp. 413–418. URL: https://search.informit.org/doi/abs/10.3316/informit.066079795299019.
[30] L. Hansen, T. Bolander, Implementing theory of mind on a robot using dynamic epistemic logic, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, International Joint Conference on Artificial Intelligence Organization, 2020, pp. 1615–1621. URL:

https://ijcai20.org/. doi:10.24963/ijcai.2020/224, twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020 ; Conference date: 07-01-2021 Through 15-01-2021.
[31] H. van Ditmarsch, W. Labuschagne, My beliefs about your beliefs: A case study in theory of mind and epistemic logic, Synthese 155 (2007) 191–209. URL: http://www.jstor.org/stable/27653487.
[32] M. Eger, C. Martens, Practical specification of belief manipulation in games, Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment 13 (2021) 30–36. URL: https://ojs.aaai.org/index.php/AIIDE/article/view/12921. doi:10.1609/aiide.v13i1.12921.
[33] A. Baltag, A logic for suspicious players: Epistemic actions and belief-updates in games, Bulletin of Economic Research 54 (2002). doi:10.1111/1467-8586.00138.
[34] Y. Qian, Y. Luo, H. Yang, S. Xie, Playing hanabi with tom and intrinsic rewards, in: PKU 22Fall Course: Cognitive Reasoning, 2022. URL: https://openreview.net/forum?id=5ckgzKj0tP.
[35] Z. Tian, S. Zou, I. Davies, T. Warr, L. Wu, H. B. Ammar, J. Wang, Learning to communicate implicitly by actions, 2019. arXiv:1810.04444.
[36] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, K. Kavukcuoglu, Asynchronous methods for deep reinforcement learning, in: M. F. Balcan, K. Q. Weinberger (Eds.), Proceedings of The 33rd International Conference on Machine Learning, volume 48 of *Proceedings of Machine Learning Research*, PMLR, New York, New York, USA, 2016, pp. 1928–1937. URL: https://proceedings.mlr.press/v48/mniha16.html.
[37] Hessel, Modayil, V. Hasselt, Schaul, Ostrovski, Dabney, Horgan, Piot, Azar, Silver, Rainbow: Combining improvements in deep reinforcement learning, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, Apr. 2018 32 (2018). URL: https://ojs.aaai.org/index.php/AAAI/article/view/11796. doi:10.1609/aaai.v32i1.11796.
[38] J. N. Foerster, F. Song, E. Hughes, N. Burch, I. Dunning, S. Whiteson, M. Botvinick, M. Bowling, Bayesian action decoder for deep multi-agent reinforcement learning, 2019. arXiv:1811.01458.
[39] L. De Raedt, A. Kimmig, H. Toivonen, Problog: A probabilistic prolog and its application in link discovery, in: International Joint Conference on Artificial Intelligence, 2007, pp. 2462–2467. URL: https://api.semanticscholar.org/CorpusID:383160.
[40] D. Fierens, G. Van den Broeck, M. Bruynooghe, L. De Raedt, Constraints for probabilistic logic programming, in: D. Roy, V. Mansinghka, N. Goodman (Eds.), Proceedings of the NIPS Probabilistic Programming Workshop, 2012, pp. 1–4. URL: http://starai.cs.ucla.edu/papers/FierensPP12.pdf.
[41] L. D. Raedt, A. Kimmig, Probabilistic programming concepts, 2013. arXiv:1312.4328.
[42] N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, I. Dunning, S. Mourad, H. Larochelle, M. G. Bellemare, M. Bowling, The hanabi challenge: A new frontier for ai research, Artificial Intelligence 280 (2020) 103216. URL: http://dx.doi.org/10.1016/j.artint.2019.103216. doi:10.1016/j.artint.2019.103216.