Hybrid Intelligence
Zero-order theory of mind

Amanda Kasat (S6153712)

February 7, 2025

1 Introduction

The Bluff game with human-agent interaction was chosen as the project. First, I wrote down the important rules of the game and then drew up a plan for how it would work.
From the game's rules, I could generate the world's rules. We have 2 players, A and B. They play using only the deck's As, Js, Qs, and Ks. In other words, 16 cards. These 16 cards are divided equally between the two players. In each round, one card type must be played: A, J, Q, and K, in that order. The player can choose to play, in which case he chooses a number N, which can be 0, in which case nothing happens, and plays N cards that are supposed to be of the corresponding type face down, or doubt his opponent's move. In this case, if the opponent's move is valid, A doubts for no reason and gets all the cards in the deck, and if the move is invalid, B gets all the cards for having "cheated."
Since it's always two players, player A knows that if he has 3 As, for example, B must necessarily have 1, i.e., the cards are complimentary. So, the strategy players can employ is to lie about the type of card they are playing. For example, player A has two As and a K, so he plays two cards in the A round, one A and one K, which can give him an advantage.
In this case, the environmental variables are:

- The number of cards in the deck
- The number of cards in the opponent's hand
- The number of cards in your hand
- The As, Js, Qs, and Ks that the opponent is supposed to have
- The As, Js, Qs, and Ks that he has

2 Zero-order model

Given the zero-order model's definition of considering environmental variables in decision-making, I need to use probability for the agents to make decisions. I've developed an algorithm that uses probability to decide when the agent finds it interesting to bluff, when it finds it interesting to be honest, or when it should doubt.

2.1 Probability Equations

Probability of doubting the opponent If the opponent has many cards of the specified type, they are less likely to be bluffing. If they don't have any cards of the type, the probability of bluffing is taken into account, given that the agent intends to get rid of their cards as quickly as possible.

$$P_{\text{doubt}} = \begin{cases} 1 - \frac{\text{cards\_player2[type\_card]}}{\sum \text{cards\_player2.values()}} & \text{if cards\_player2[type\_card]} > 0, \\ 1 & \text{otherwise.} \end{cases}$$

Probability of bluffing When the number of cards played (n) exceeds the number of cards available in the type`cards_player1[type_card]`, the probability of bluffing increases proportionally. If the player has few cards in total (< 4), he can bluff more often (minimum probability of 50%. Otherwise, he doesn't bluff.

$$P_{\text{bluff}}(n) = \begin{cases} \frac{n - \texttt{cards\_player1[type\_card]}}{n}, & \text{if } n > \texttt{cards\_player1[type\_card]}, \\ 0.5, & \text{if } \sum \texttt{cards\_player1.values()} < 4 \text{ and } n \leq \texttt{cards\_player1[type\_card]}, \\ 0, & \text{otherwise.} \end{cases}$$

Probability of which card will be chosen when bluffing The probability of each card is calculated in proportion to the number of cards of that type in the player's deck and their weight, whereby cards with a greater product between quantity and weight are more likely to be chosen. Cards that would only be played in more distant rounds have more weight.
The probability of choosing a card c is given by:

$$P(c) = \frac{\text{cards\_player}[c] \times \text{weights}[c]}{\sum_{c' \in \{A,J,Q,K\}} \text{cards\_player}[c'] \times \text{weights}[c']}$$

Where:

- cards_player[c] is the number of cards of type c the player has.
- weights[c] is the weight assigned to the type of card c (usually calculated by an external function like `updates_weights`).

Expected value for each action The player loses the total number of cards in the deck ($-\text{len(deck)}$) if they are caught bluffing ($P_{\text{bluff}}(n)$). If the opponent does not doubt ($1 - P_{\text{doubt}}$), the player incurs no penalty (0).

$$V_{\text{honest}} = P_{\text{bluff}}(n) \cdot (-\text{len(deck)}) + (1 - P_{\text{doubt}}) \cdot 0$$

If the opponent doubts (P_{doubt}), the player loses the cards in the deck. If the player bluffs ($P_{\text{bluff}}(n)$), they may also be penalized. In the case where the opponent does not doubt and the bluff succeeds ($1 - P_{\text{doubt}} - P_{\text{bluff}}(n)$), the player loses the number of cards they played (n).

$$V_{\text{bluff}} = \max\left(P_{\text{doubt}} \cdot (-\text{len(deck)}) + P_{\text{bluff}}(n) \cdot (-\text{len(deck)}) + (1 - P_{\text{doubt}} - P_{\text{bluff}}(n)) \cdot n\right)$$

If the opponent is honest, doubting causes the player to lose all the cards in the deck. If the opponent is bluffing, doubting causes the player to gain all the cards in the deck.

$$V_{\text{doubt}} = \begin{cases} -\text{len(deck)}, & \text{if the opponent's move is valid,} \\ +\text{len(deck)}, & \text{if the opponent's move is invalid.} \end{cases}$$

Final decision The player chooses the action that maximizes the expected value. In other words, he moves with the greatest benefit or the least risk.

$$\widehat{\text{Decision}} = \arg\max \left(V_{\text{honest}}, V_{\text{bluff}}, V_{\text{doubt}} \right)$$

3 First-order model

Unlike the zero-order model, the first-order model must think about what the other player is considering to make its decisions. To do this, the player needs to analyse both the past move, i.e., whether the opponent bluffed or played honestly, and the prediction of the next move, i.e., whether their opponent will bluff, doubt, or play honestly on their next move.

3.1 Prediction of opponent's behavior

Initially, the probabilities of the opponent being honest (Phonest), bluffing (Pbluff), or doubting (Pdoubt) are set to initial values of 0.45, 0.45, and 0.1, respectively. These values were chosen based on the assumption that in an average situation the opponent has a reasonably equal chance of being honest or bluffing, with a lower probability of doubting.
We then adjust the odds according to the estimated cards of the opponent's current suit. In this context, if the opponent has many cards of the current suit, the chances of him bluffing are high. In addition, if the player thinks that his opponent bluffs in the previous round, the probability of him bluffing again in this round also increases. In addition, if the player has few cards in his hand and plays a lot of cards in the round, the likelihood that his opponent doubts is high.
Finally, a normalization is made that ensures that the Phonest, Pbluff, and Pdoubt probabilities add up to 1, keeping them comparable and meaningful. This avoids distortions and ensures that the agent's decisions are based on a valid probability distribution.

3.2 Equations

Probability of being honest Vhonest increases if the opponent has few cards or if the probability that the opponent doubts is high, because this would make it more likely that the player will decide to be honest, since the opponent may doubt or not have many cards to support a bluff.

$$V_{\text{honest}} = (1 + P_{\text{doubt}}) \cdot n - (1 - P_{\text{doubt}}) \cdot \text{len(deck)}$$

Probability of bluffing Vbluff is calculated by considering the following situations: first, the case where the opponent doubts the move, which results in a penalty based on the deck size; second, the case where the opponent believes the bluff, where the expected value is based on the number of cards played; and third, the case where neither doubt nor belief happens, with the value being a combination of these scenarios weighted by the probabilities of doubt and bluff.

$$V_{\text{bluff}} = \max \Big(P_{\text{doubt}} \cdot (-\text{len(deck)}) + P_{\text{bluff}} \cdot (-\text{len(deck)}) +$$
$$(1 - P_{\text{doubt}} - P_{\text{bluff}}) \cdot n \,\Big|\, n \in \{1, \ldots, \text{cards\_player1[type\_card]} + 1\} \Big)$$

Probability of doubting If the opponent has made an "H" (Honest) or "B" (Bluffing) move, the function checks whether the opponent has few cards (3 or less) and has already played 2 or more cards. If these conditions are true, the probability that the opponent is bluffing increases, and therefore the advantage of doubting (Vdoubt) is adjusted, otherwise it is not advantageous.

$$V_{\text{doubt}} = \begin{cases} P_{\text{doubt}} & \text{if opponent\_hand\_count} \leq 3 \text{ and } n_{\text{played}} \geq 2 \\ -\text{len(deck)} & \text{otherwise} \\ -\infty & \text{if the player started the round} \end{cases}$$

4 Observed results

First, I pitted two zero-order agents against each other. I repeated the experiment 25 times, in which player 5 won twice, as being honest was a more interesting strategy [2] and, on average, the match lasted 10 rounds [1]. However, in most cases there is a "draw", which makes sense because if both players have the same strategy of considering only the environmental variables when making decisions, the most determining factor for one of them to win is the initial distribution of the cards. Otherwise, they will remain in a loop and more doubt actions happen.

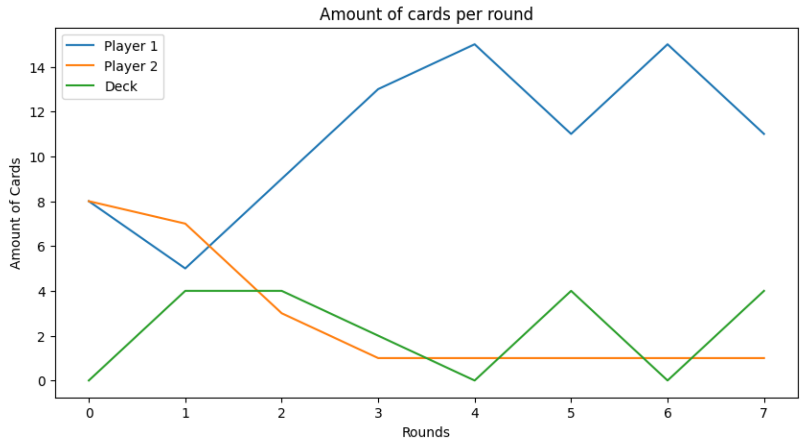

Figure 1: Distribution of the cards in the game in a case where player 2 won.

After that, I repeated the experiment, but this time with the first-order agent against the zero-order agent. In this situation, the matches lasted an average of 14 rounds, with the first-model order winning almost all of them. The match ends up being longer than the zero-order versus zero-order match, because in this case the first-model tries to predict the opponent's behavior, which leads to more doubting and bluffing, which increases the number of rounds if the player is penalized. This increase in actions other than honesty [4], with bluffing being the action most often taken, also causes the number of cards in each player's hand to vary more, making the graph more unstable [3].

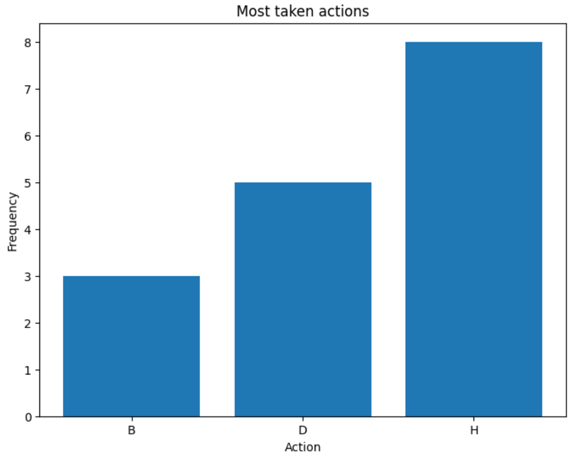

Figure 2: Distribution of actions taken in the game in a case where player 2 won.

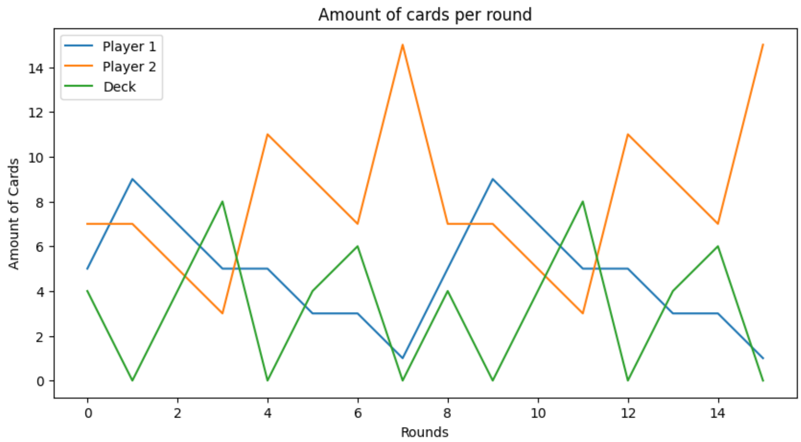

Figure 3: Distribution of the cards in the game zero-order vs. first-order.

I also tested the first-order model against itself. In this case, the number of rounds dropped by half because both try to predict each other's behavior, resulting in better mutual understanding and more efficient strategies. This reduces ineffective actions, such as unsuccessful bluffs or unnecessary doubts, allowing the game to be resolved more quickly. Bluffing continued to be the most used strategy, which also explains the reduction in games, and few doubtful actions were taken. The graph ended up being more stable than in the first-model vs. first-model comparison because the players follow more predictable and consistent behavior patterns.

Finally, we tested each model with humans, asking five people to play five games against the zero-order and five against the first-model. In this way, I obtained 25 data points on matches against the zero-order model and the same amount for the first-order model. Of the games played, the zero-order won only 3 out of 25 times, and the first-order won 7 out of 25.

According to the reports of the people who played against the zero-order model, they noticed that, almost always, at the end of the game, if they played a lot of cards, the model was suspicious,

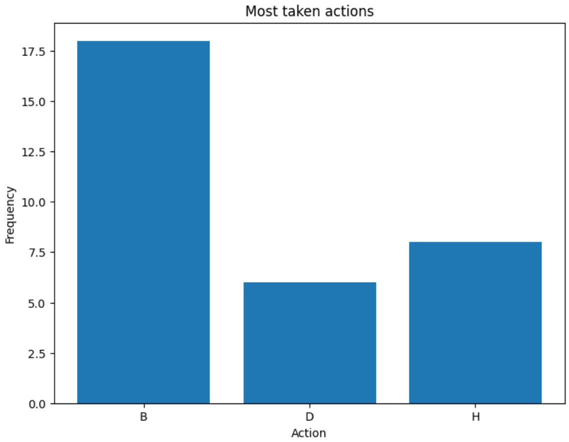

Figure 4: Distribution of actions taken in the game zero-order vs. first-order.

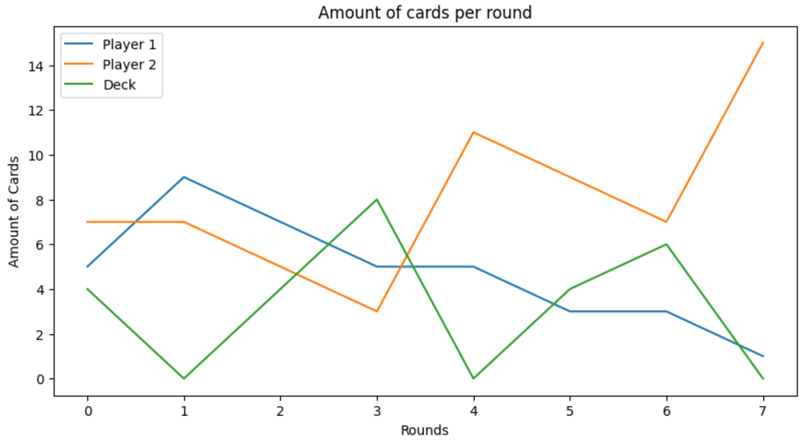

Figure 5: Distribution of the cards in the game first-order vs. first-order.

whereas if they played a lot at the beginning, the doubting action happened much less. In this way, playing several cards at the start of the game, especially those of the suit furthest away at the moment, turns out to be advantageous.

In general, the matches had 4 to 5 rounds. In the graph [7] we can see that in the third round the number of cards in the agent's hand goes up, this is due to that doubting action at the end of the cycle which often went wrong for the agent. You can also see that when the game went beyond 5 rounds, it was much more likely that the agent would win, because for the game to have gone on like that, it took a bluff or a poorly executed doubting action on the part of the human player, which left a lot of cards in the player's hand, leaving him at a disadvantage.

Compared to the results between human vs. zero-order and human vs. first-order, the graph is not so different from each other. However, something important is that at the end of the graph [8], we can see that the first-order is left with several cards, because he tries to predict that the player will bluff, because he is at the end of the game, but if the player is honest, this strategy is not good.

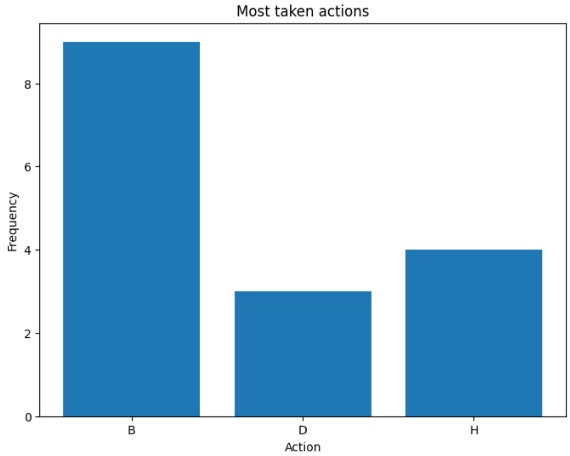

Figure 6: Distribution of actions taken in the game first-order vs. first-order.

According to reports from people who have played against the model, the first-order model is much more suspicious of game actions. So bluffing attempts are more likely to be caught, and playing honestly is an interesting strategy that also greatly reduces the number of moves. Because the agent will think that the other is bluffing, and end up keeping the cards in the deck for himself.
Thus, when tested against humans, the results indicate that honesty can be an effective strategy against both models, but especially against order one, since bluffing attempts are caught more often. In other words, bluffing is more interesting with the zero-order model, as it is also much more predictable for human players, as it follows a more visible pattern of behavior.

5 Conclusion

In this project, I developed and analyzed a human-agent interaction model for the game Bluff. We used two types of agent: a zero-order agent, which only considers environmental variables, and a first-order agent, which tries to predict the opponent's behavior based on its previous actions.
The experiments showed that the first-order agent outperformed the zero-order agent, both in matches against other agents and in interactions with humans. The ability to predict the opponent's behavior provided the first-order agent with a more efficient strategy, resulting in longer and more dynamic matches, with greater variation in bluffing and doubting actions.
Analysis of the human players' reports showed that it was easier to exploit the zero-order agent's behavior patterns, while they had greater difficulty anticipating the first-order agent's strategies.

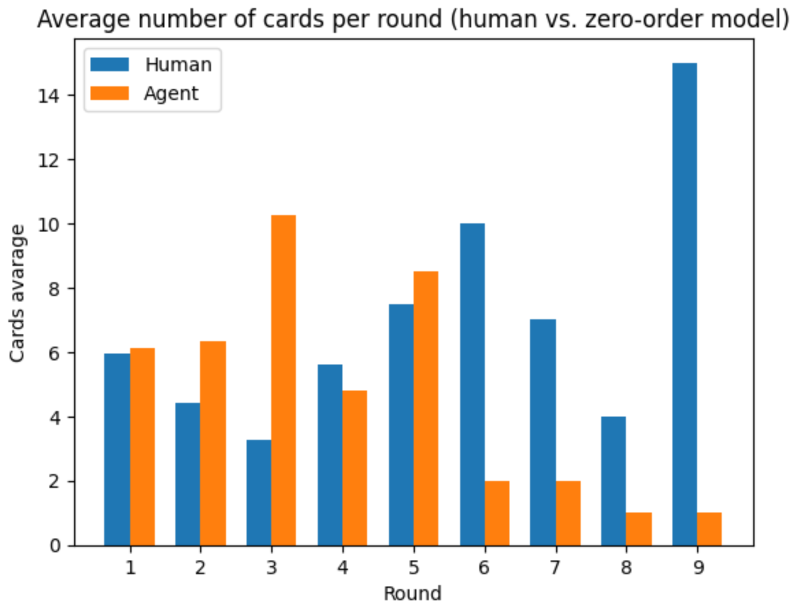

Figure 7: Distribution of cards in the game human vs. zero-order.

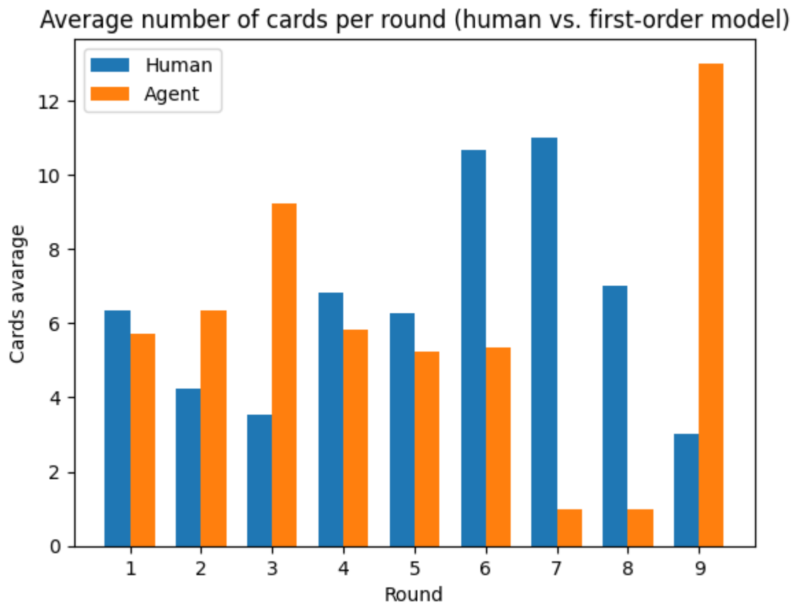

Figure 8: Distribution of cards in the game human vs. first-order.