# Integrating fairness in opponent modeling

## Problem

In an automated negotiation the common goal is to achieve the best outcome possible for the parties involved. What does best mean in this context?

- 1. There are different metrics in terms of which we can say if an outcome is good or not
- 2. An intuitive response: best == fair

In automated negotiation there are three main components: Bidding strategy, Acceptance Strategy and Opponent Modeling a.k.a BOA framework [3]. The opponent modeling plays an important role:

- 1. Due to agents keeping their preference profile private
- 2. Competitive agents
- 3. High bid spaces

It is then reasonable to investigate how can fairness be achieved in opponent modeling, as it affects the outcome of a negotiation:

- 1. Research question: Is there a way to integrate fairness by way of opponent modeling?
- 2. More specifically: How fair is the Bayesian Opponent model strategy?

## Strategy Analysis

The analysis of the Bayesian learning technique revealed key component, that has a direct impact on the efficiency of the opponent model, and consequently on the negotiation.

### Assumptions

- Structural assumptions
- Rational assumptions

## Related work

- Many research attempts to define fairness and its application in computer science[1]
- Research in how to measure fairness [2]
- Research in different opponent models

What about fairness in opponent modeling?



## Fairness

Questioning what is fairness is a result of "centuries of oppression of people based on race, gender, and social class", and looking for answers in philosophy leads to useful reference points.

### Why is Rawlsian approach useful?

Fair way to distribute goods among people with distinct morals

The Difference principle - distribution is in favour of the least advantaged

Objective decisions Veil of Ignorance, disregard sensitive inequalities (race, gender, social class)

#### Procedural fairness

Study fairness in the process of negotiation

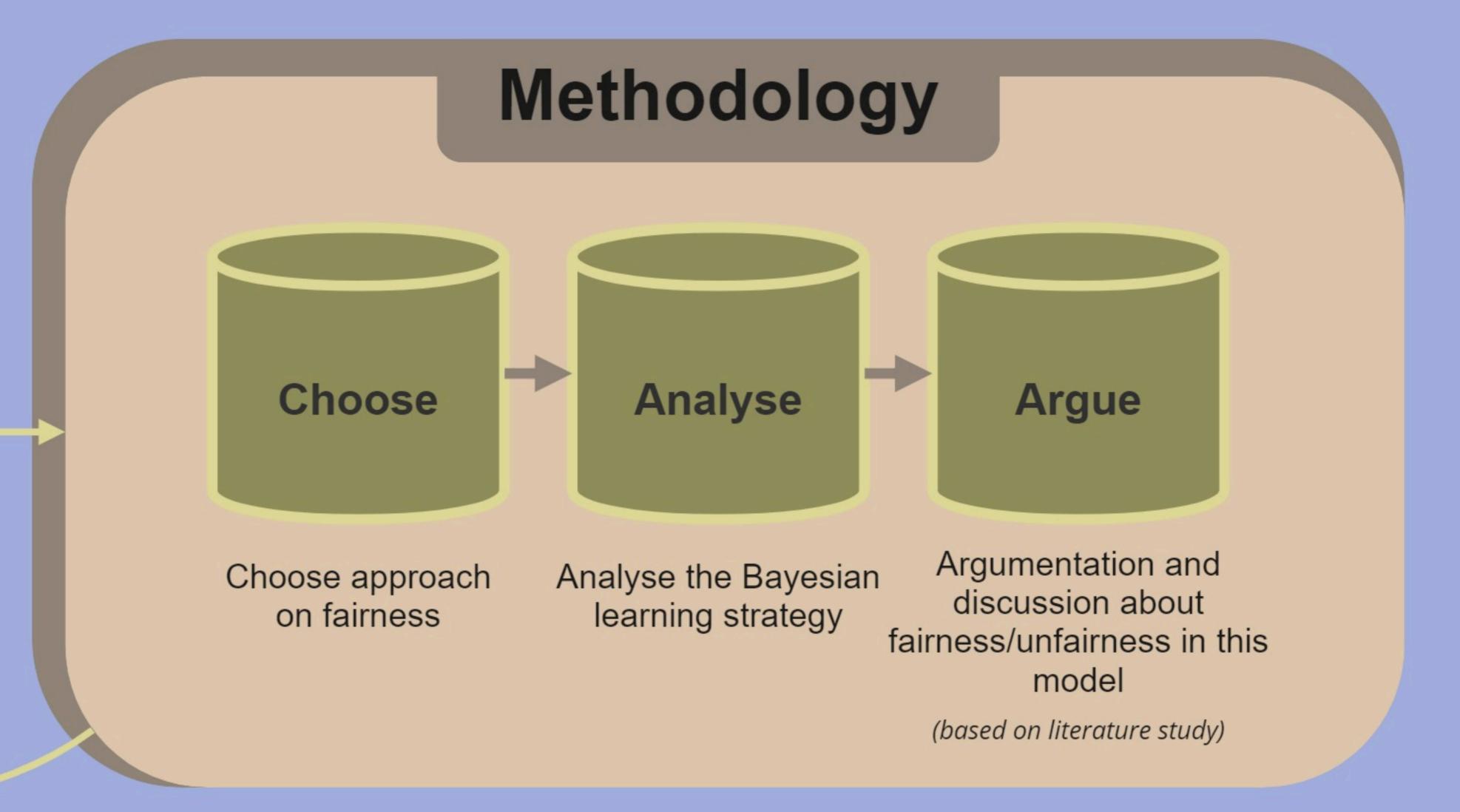
## Unfairness

### Why use these assumptions?

- 1. Give meaning to the observed
- 2. Without them the process can become ineffective, and lead to unwanted results

## What is the problem then?

- 1. These assumptions impose hard constraints on the model - agent cannot negotiate optimally with irrational opponents
- 2. Bias The agent has a biased behaviour towards players who follow concessionbased and time-dependent tactics.
- 3. The Bayesian model does not guarantee a successful estimation of the opponent's preference profile because of the rational assumptions



## Conclusion

- Opponent modeling plays an important role in automated negotiation
- Assumptions are harmful in some contexts
- Challenge towards developing a fair process in a negotiation
- Bias can be present in opponent modeling and we need to mitigate this
- The Bayesian model hides unfair aspects that lead to an overall unfair negotiation

## Future work

Analyse fairness in other opponent models Adapt or create opponent models that are tailored towards incorporating fairness

Create a general framework that measures fairness in the process of opponent modeling



[1] Sahil Verma and Julia Rubin. Fairness definitions explained. International Workshop on Software Fairness, 2018.

[2] Z. Jacobs Abigail and Hanna Wallach. Measurement and fairness. FAccT 21 Virtual Event, Canada, 2021.

[3] Baarslag, T., Hendrikx, M. J., Hindriks, K. V., & Jonker, C. M. (2016). A survey of opponent modeling techniques in automated negotiation.