

The Unintended Consequences Fairness Brings to Automated Negotiation

Improving fairness in automated negotiation

1) Introduction

Automated negotiation is becoming more integrated into everyday life, e.g. e-commerce and resource allocation. To combat parties from pushing for their own gain and for automated negotiators to properly represent their user [1], having some notion of fairness applied to the negotiation is crucial. However, implementing fairness into the automated negotiation process might introduce unintended consequences.

2) Relevance

Unintended consequences could be, in the worst case, vulnerabilities that when not taken care of could lead to exploitation by a malicious party for their personal gain, contradicting the general idea of fairness in automated negotiation and often essentially rendering the implementation of fairness useless.

3) Justice as fairness

'Justice as fairness' is John Rawls' take on the definition of fairness. In *Justice as Fairness: A Restatement* [2]. Rawls describes his idea on fairness, which consists of two principles:

1. **The liberty principle**
2. **The equality principle**

The second principle, which indicates when inequalities are allowed, consists of two parts:

1. **Fair equality of opportunity**
2. **Difference principle**

Liberty principle

every person has the same baseline level of liberties and freedoms. The liberties of certain people or groups could be toned down for the sake of overall liberty.

Fair equality of opportunity

Any person should have a fair chance to obtain benefits in society.

Difference principle

Inequalities should always be to the benefit of the least advantaged. The advantaged group should not be able to gain more at the expense of the lesser advantaged.

When using a Rawlsian notion of fairness in automated negotiation, are there any unintended consequences?

4) Method

An automated negotiation process is setup with two agents. Agent 1 is considered the advantaged group, agent 2 the disadvantaged.

- **Bidding Strategy:** Trade-off strategy
- **Acceptance Strategy:** Utility of incoming bid is larger than utility of the agent's counterbid
- **Opponent modelling:** Agents know each others preference profiles

During the negotiation, agent 1 cannot decrease the utility of agent 2 for the bids its sending. This imitated the difference principle, the idea that the advantaged should not benefit at the expense of the disadvantaged.

Multiple domains are checked for unintended consequences. An unintended consequence is any peculiar behaviour in the negotiation process or an outcome that is not fair in accordance to Rawls' justice as fairness.

5) Results

- The bidding space of agent 1 has shrunk because it can only increase opponents' utility, leading to many of the same bids over a period of time.
- The outcomes of the negotiations were not always optimal (on the Pareto-front), thus did not conform to the difference principle by default.
- No unintended consequences that can be proven to be strictly related to Rawls' notion of fairness were found.

6) Conclusion

The experiment showed some unharmed unintended consequences, however they seem to be context- and implementation-specific, not necessarily purely related to Rawls' notion of fairness. However, this does not mean that justice as fairness has no consequences. Since the unintended consequences seem context-related, this is a good pointer for further research.

7) Further Research

Further research on finding (other) unintended consequences include:

- Test other notions of fairness
- Improve operationalization of Rawls' notion of fairness
- Change or expand on what an unintended consequence defines
- Put agents more in context of society

8) References

[1] T. Baarslag, M. Kaisers, E. H. Gerding, C. M. Jonker, and J. Gratch, "When Will Negotiation Agents Be Able to Represent Us? The Challenges and Opportunities for Autonomous Negotiators," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, (Melbourne, Australia), pp. 4684–4690, International Joint Conferences on Artificial Intelligence Organization, Aug. 2017.

[2] J. Rawls and E. Kelly, *Justice as Fairness: A Restatement*, Cambridge, Mass: Harvard University Press, 2001.