Investigation of the fairness metrics in automated negotiations

Andrzej Rubio Bizcaino
Supervisor: Sietze Kuilman
Responsible Professor: Luciano Cavalcante Siebert
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

**A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering**

Abstract

This paper aims to define the broad concept of fairness and investigate how it can be measured, especially considering fairness in automated negotiations. The report relies on the work on fairness issues that have been derived from the research of C. Albin [1]. Firstly, the paper elaborates on different fairness metrics from the literature review. Then these metrics are tested to assess if they capture the effect of considering fairness by the agents in the negotiations. That is done by simulating multiple bilateral negotiations under the open-source *GeniusWeb* framework, where eccentric agents are compared with fairness-oriented parties by using the fairness metrics. Based on the conducted experiment, the most consistent fairness metric in automated negotiations is the distance to the game-theoretic solutions to the bargaining problem, which considers much of the outcome fairness concept. However, other investigated metrics also capture the different scope of fairness and can be used as metrics, especially when combined, and some interdependence between metrics is modelled.

1 Introduction

Various types of negotiations are omnipresent in many areas of everyday life, ranging from the negotiations concerning the pricing of goods, employment, loan conditions and many more areas varying from commerce through politics to science and technology. However, T. Sandholm observes that because of recent advancements in Artificial Intelligence, there is a shift from human agents taking part in the negotiations to fully computational agents participating in the negotiations on behalf of the parties they represent [20]. Although some researchers argue that we should tend to integrate AI with human agents in the negotiation [14], T.Sandholm draws many advantages of purely automated negotiations. He claims that the time a human party needs to be attentive is radically shortened, the negotiations are free of emotions and external bias, which may lead to irrational behaviour, and a reasonable agreement settlement can be found even in highly complex negotiation settings, where negotiations concern many issues [20]. Although the automation of bargaining can lead to many improvements, the exact implementation of the negotiation agents becomes a fundamental issue. In the context of automated bargaining, it is essential that the agent not only maximize individual utility but also lead to a situation where the negotiation can be regarded as fair to all participating parties [23].

However, the idea of fairness is a highly comprehensive concept, which many great philosophers and psychologists have been trying to define throughout the ages [17]. The first idea of fairness is dated back to Ancient Greece, where philosophers tried to prove the benefits of just behaviour and introduced some first takes on looking at fairness. As an example, Socrates regarded fairness as the sense of duty to the society, and Aristotle incorporated the idea of proportionality, which in the context of fair allocations claims that there must be proportionality between the input the party brings to the society and the output it receives from it [17]. An instance of one of the contemporary discussions concerning fairness could be the debate between American philosophers John Rawls and Robert Nozick [24]. The views of Rawls are much similar to the ones presented by the egalitarian movement, which proposed that all should be able to have equal access to specific opportunities and that we should strive for a situation where there is no inequality, and if that is impossible, then at least try to make those who are least advantaged be best off [25]. Rawls also introduced the concept of 'Veil of Ignorance', a powerful philosophical concept which implied that every party should assess the position of other parties and try to equalize them so that no one would feel disadvantaged if one were represented by the randomly chosen party [11]. Rawlsian ideas conflicted much with Nozick's take on fairness, who argued that it is, in essence, unfair to alter parties' position in a status-quo "just situation" to compensate for the situation of other parties. On the other hand, the philosophical view of communitarianism regarded that fairness should be mainly viewed from the perspective of a community rather than an individual [10].

These many perspectives on fairness put high importance on the precise formulation of the definition of fairness, especially in the context of automated negotiations where agents should follow the just principles. Unfortunately, because of many different and contradictory takes on fairness, few research papers define a general framework on fairness for agents participating in negotiations, and there are limited ways to measure fairness, especially in automated negotiations. Thus, the main aim of the research paper is to investigate possible ways to quantify fairness in the context of automated negotiation and consider which ones are best suited for this problem. For that purpose, this paper starts by describing different notions of fairness incorporated into the strategy of computational agents developed with the open-source GeniusWeb framework [16]. Then the metrics found in the literature are assessed by running multiple negotiation sessions and comparing the difference in their assessment between eccentric and developed fairness-oriented agents.

In this paper, I investigate metrics describing the vast concept of fairness within automated negotiations under the following structure. First, in Section 2 I provide some background to different notions of fairness in negotiations, considering which areas the concept of fairness can be contested. That leads to the precise problem formalization supported by the research on this topic's literature. Then, Section 3 elaborates on the conducted empirical research setup. Additionally, that section briefly presents the assessment methodology for the derived metrics. Next, in Section 4, I deliver the results of the investigation on the performance of fairness metrics. Then, I discuss the results in Section 5, with an indication of which metrics captured different notions of fairness and, based on that, suggest which metrics are well suited for the problem of automated negotiations. Additionally, some general reflections on the conducted research are included to support the potential further improvements to the study to validate the findings. Section 6 briefly elaborates on the considered issues regarding the responsible research. Lastly, in Section 7 I summarize my investigation and conclude with some general remarks about the topic of fairness measurement in automated negotiations.

2 Background & Problem Formalization

Because of the broad scope of fairness, it is essential to base the research on a framework which would define the fields in which one can investigate fairness. Furthermore, that framework on fairness issues is critical when considering the fairness-oriented agents, as such agents should incorporate 'fair behaviour' concerning at least one of the described issues. Additionally, this framework will be crucial in considering the fairness metrics, as in a perfect setup, they should also be sensitive to the issues of fairness in negotiation. Therefore, I focus on the research with the framework of fairness issues in negotiations conducted by C. Albin [1], as she is one of a few who dedicated a paper on that topic with such an extensive exploration of areas on which researchers can contest fairness. These fairness issues are discussed in more detail in the following Section 2.1.

2.1 Fairness Issues In Negotiation

C. Albin argued that the concept of fairness within the negotiations could be categorized into four different issue groups of: **Structural Fairness**, **Process Fairness**, **Procedural Fairness** and **Outcome Fairness** [1].

The class of **Structural Fairness** considers fairness from the perspective of the overall structure of the dispute and the relations between parties and negotiated issues. The key components that consider Structural Fairness are the overall physical, social, and issue constraints for which the negotiation proceeds and negotiators operate. For example, Structural Fairness will try to answer if the negotiation protocol is not discriminating against any party and if all participating parties have a genuine opportunity to propose all possible bids within the negotiation. Although extremely crucial to fulfilling, this class of fairness issue could be vague to incorporate into the general behaviour of the negotiating parties. Thus most of the issues considered by the Structural Fairness will be maintained by the general setup of the conducted negotiation simulation, described in Section 3.

Process Fairness considers how the opposing parties relate to each other and how parties' ideas of fairness influence the dynamics of the negotiation process. This area, in essence, describes the similarity between the parties and just behaviour as a response to the way the opposing party takes part in the negotiation. An example of the issue belonging to the Process Fairness class is whether it is fair to not bother with the opponent's utility from the final settlement if the opposing party is proposing only self-beneficial offers and is unwilling to make any concessions.

The next category of fairness issues is the class of **Procedural Fairness**, which regards the specific mechanism and strategies that govern a party and whether they can be considered fair because of some intrinsic value and because they lead to fair outcomes. By way of illustration, an issue belonging to Procedural Fairness would be much concerned about the exact mechanism, like a question if a party should make 'equal' concessions in negotiations as its opponent. The category of Procedural Fairness may sound familiar to the previously described class of Process Fairness. However, Procedural Fairness considers how a party's specific strategies are implemented during the negotiation, whereas Process Fairness regards the change between strategies according to the opponent's behaviour.

That brings us to the last group of fairness issues described by C. Albin, namely the **Outcome Fairness**. It refers to the specific principles concerning the fair outcome from the final negotiation settlement. Some prominently described principles regarding the final allocation are equity, equality, and need [7]. Briefly elaborating, the equity principle mirrors Aristotle's idea of fairness through proportionality between input and output. The equality principle assumes that the parties should receive the same or similar rewards from the negotiation, respectively, to their situation. Finally, the need principle states that the resources should be allocated according to the intrinsic necessity so that the least advantaged agents receive the most significant share. As seen, the different notions governing Outcome Fairness can be exclusive. The ideas of Outcome Fairness are taken into account when incorporating metrics related to the solutions of the renowned, game-theoretic bargaining problem defined by J. Nash [15].

2.2 Fairness Metrics

As said in the Introduction, there is no generally developed framework for measuring different notions of fairness in the negotiations, posing a significant problem, especially for automated negotiations. The inexistence of a specific framework is partially understandable as using specific metrics, one indirectly incorporates own perspective on fairness. This subsection elaborates on some most prominent fairness metrics found in literature, which are assessed in the experiment described later in the paper.

Distance to Bargaining Problem Solutions

The foremost considered metric to assess fairness is the distance to the proposed solutions to the bargaining problem described by J. Nash [15]. The bargaining problem is a mathematical representation of the negotiations, where individuals with opposing interests need to agree on the final allocation. Thus, the solutions Bargaining Problem pose an analogy to the Outcome Fairness described by C. Albin. The individuals act according to their utility function $u(\cdot)$, which maps offered bids to the real number representing how beneficial a bid is for the party. A key concept within the topic of the bargaining problem (and negotiations in general) is the Pareto Frontier - a set of all solutions such that no individual can be better off without making at least one opponent worse.

The first solution is a Nash Product solution proposed by J. Nash [15], which is a Pareto optimal point that maximizes the product of utilities of the participating parties. This solution seeks the most optimal solution for all parties, with penalizing scenarios in which there is high inequality between the obtained utilities by the parties. Under axioms derived in his paper, J. Nash argues that this solution always yields a fair outcome under different variations of the utility space.

The next considered solution to the bargaining problem is a Kalai-Smodrinsky solution, which seeks a' fair' solution for negotiation parties by equalizing the ratios of obtained utilities to the utility obtained when no agreement is reached (reservation value) [13]. When both parties have the same reservation value, the Kalai-Smodrinsky solution is a Pareto optimal point closest to the $x = y$ line where the axes reflect the utilities of the participating agents. In a Figure 1 the toy scenario of the bargaining problem has been represented with a visualization of both Nash Product and Kalai-Smodrinsky solution.

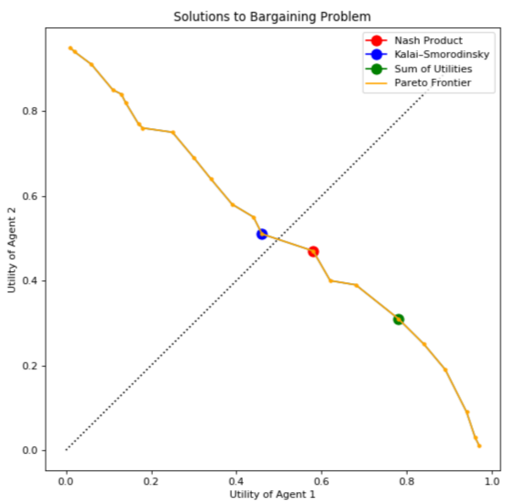

Figure 1: Representation of the Pareto Frontier with marked Kalai-Smodrinsky & Nash solutions to the bargaining problem, as well as bid representing maximal sum of individual utilities

Sum of Individual Utilities

An additional solution represented in Figure 1 is the point representing the sum of the individual utilities [9]. Like the Nash Product, that solution to the Bargaining problem seeks an overall most profitable final allocation without penalizing the inequalities between the parties' final utilities. Therefore, it fits the more individualistic notion of fairness as although it tends to the highest aggregate utility, it will also tend to favour parties that can obtain most of the negotiation. In this research, the distance to the maximum sum of individual utilities and the final sum of individual utilities are considered fairness metrics.

Kindness Function

The next investigated fairness metric is a Kindness Function derived by the M. Rabin [18]. As a cornerstone of his research, he derives three insights regarding the negotiations:

- People are willing to sacrifice their well-being to help those who have been kind
- People are willing to sacrifice their well-being to punish those who have been unkind
- Both motivations have a greater effect when the cost of sacrifice becomes smaller

Based on these observations, he defines the Kindness Function as a normalized notion of how beneficial bids are for an opponent from equitable payoff, assuming that parties act according to concrete negotiation strategies. For his implementation of the kindness function, he defines a_i as a known strategy of party i, b_j as an assumed strategy of party j (describing how the opposing party, e.g. i, thinks a strategy of party j is) and utility function $u_j(b_j, a_i)$ as a utility of opposing party j assuming it follows strategy b_j and that self party is following known strategy a_i. He also defines $u_i^h(b_i)$ as a highest utility payoff of party i assuming it follows strategy b_i, and $u_i^{min}(b_i)$ as a lowest utility payoff of party i assuming it follows strategy b_i. Lastly he defines $u_i^l(b_i)$ as party i lowest utility among points that are Pareto-efficient and $u_i^e(b_i) = [u_i^h(b_i) + u_i^l(b_i)]/2$ as a equitable utility. If Pareto Frontier is linear, the equitable utility is a middle point between most outer Pareto optimal solutions. In the Equation 1 the precise calculation of the Kindness Function representing how kind a party 1 is towards party 2 is included.

$$f_1(a_1, b_2) = \frac{u_2(b_2, a_1) - u_2^e(b_2)}{u_2^h(b_2) - u_2^{min}(b_2)} \tag{1}$$

Equation 1: Kindness Function. The $u_2(b_2, a_1)$ represents the utility of the party 2 for the proposed bid by party 1

Within the research, I look at the metrics from a static perspective and not as an indicator of how parties should change their behaviour. Additionally, some investigated agents do not use opponent strategy modelling. Thus, I use the simplified version of Kindness Function, which concerns normalized distance between the utility of the proposed bid to the equitable utility defined by M.Rabin.

Fluctuations in final agreement

The next considered metric is the final agreement fluctuations across various negotiation simulations. According to C. Dwork et al. [8], similar parties should be treated similarly across many held negotiations. Thus, I use the sum standard deviation between utilities of participating parties from the utilities of the final allocation gathered from multiple negotiation simulations. This metric considers fairness in negotiations from the perspective of loss minimization problem as the metric would reward parties that endeavour to remain consistent in their final allocations if they negotiate with similar parties.

Time to an Agreement

Although the elapsed time to an agreement is not a metric that can be directly related to fairness, some researchers argue that the elapsed negotiation time can be a straightforward measurement of the performance of the different strategies from various perspectives [19]. It also measures the efficiency of the computational power of the negotiating agents, which could be argued to be a desirable quality for the parties taking part in the negotiations. However, from the preliminary research, it may seem that this metric would hardly capture the fairness consideration within the agents.

3 Simulation Setup

To investigate which fairness metrics are best suited for the automated negotiations, fairness-oriented and eccentric automated agents have been put under multiple negotiations and assessed by the fairness metrics. Then the difference in metrics values assessing eccentric and fairness-oriented agents are compared. Thus, this section defines the precise negotiations' setup.

3.1 Negotiation Protocol

The negotiations follow the Stacked Alternating Offers Protocol (SAOP) negotiating protocol [2]. The protocol takes the turn-taking order until reaching a consensus or a termination condition, after which both parties obtain their reservation value. The reservation value for both parties is set to 0. For each turn, a negotiating party that received an offer can either:

- accept the obtained offer proposed by the opponent
- make a counteroffer and propose it to the opponent
- end negotiation

At the start of the negotiation, the first party can either make an offer or end the negotiations. In practice, because of the non-positive reservation value combined with the specific bid-space domain, which always yields positive utility values (thoroughly described in Section 3.2), the negotiating parties will always be better off when accepting even the least beneficial bid than to end the negotiations. The negotiations end once they reach the end of a 200th round. Each round includes two turns, one per negotiating party, so the maximum combined number of offers is 400.

3.2 Domain & Utility Space

The used domain in the negotiation simulations is the *7issues* domain available in the open-source *GeniusWeb* framework [16]. This domain is a most extensive bid-space domain from the one available in the core *GeniusWeb* framework. It consists of the seven negotiation issues, each having ten possible negotiation values. That results in ten million possible bids, so not all bids may be offered during a single negotiation.

The utility spaces of the participating agents follow the *Linear Additive Model*. Each issue has an assigned weight, as well as each value of that issue. Thus, the utility of a bid is the sum of the issues' weights multiplied by the weight of a related issue's value included in the proposed bid. This can be represented by equation $U(o) = \sum_{i \in Issues} w_i \cdot V_i(o_i)$, where w_i represents the weight of the issue i, and $V_i(o_i)$ represents weight of a value from bid o for issue i. The utilities are scaled to fit a $[0, 1]$ range. The visualization of all possible bids on the utility space of the negotiating parties is shown in the Figure 2, with marked solutions to the bargaining problem for that utility space and the point representing a maximal sum of individual utilities.

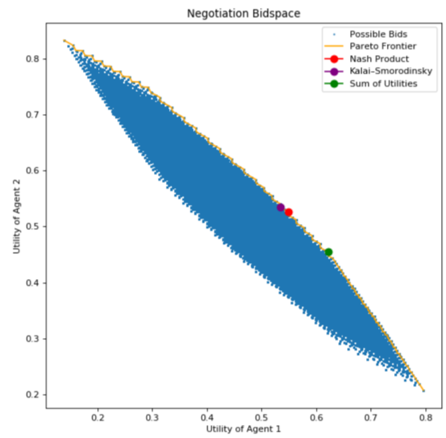

Figure 2: Bid-space of the negotiation, representing utilities for the participating negotiating agents for the all possible bid outcomes

Because of an extensive number of bids, the bid space forms an almost continuous convex set. That is the desired quality as it offers many variations in possibly offered bids, and because of its' convex-like shape, it prevents any edge-case behaviour of the investigated agents.

3.3 Opponent Modeling

As some investigated agents need to model the preferences of an opponent, it is crucial to develop a shared opponent modelling to maintain all factors equal in the experiment setup. As the Linear Additive domain models the agents' utility, the opponent modelling considers estimation of both issue weights and corresponding values described in the following subsections.

Issue weight estimation

The opponent is unlikely to concede on issues most important to them. Thus, the more important an issue is, the less likely it is that the opponent frequently changes offered values [5]. Therefore, the developed issue weight estimation is divided into two phases:

1. Updating weights with learning rate (α = 0.1) - The opponent's last two offered bids are compared and checked if the offered value has changed. If not, then the weight of an issue is increased by value alpha $w_i = w_i + \alpha$, and if the values change, the weight is not updated.
2. Normalization - As, the issue weights have to sum up to 1 $\sum_{i \in Issue} w_i = 1$, the weight of each issue is divided by their sum $w_i = \frac{w_i}{\sum_{i \in Issues} w_i}$. With that, the less diverse issues increase in their weight, whereas the frequently-changing issues have their weight decreased.

Values' weights estimation

Similarly to issues' weights estimation, the values' weights estimation uses a frequency calculation model, which increases the value of the most frequently used values. In addition, the model uses Laplace smoothing of the utilities [21]. With that, one gets the value weight estimation obeying the following Equation 2.

$$V_i(o_i) = \frac{(1 + \sum_{j=1}^{n} \delta_j(o_i))^{\gamma}}{max_{i \in Issues}(1 + \sum_{j=1}^{n} \delta_j(o_i))^{\gamma}} \quad (2)$$

Equation 2: Function representing value weight estimation

Where $\delta_j(o_i) = 1$ if value o_i has appeared in bid j. Otherwise the value of $\delta_j(o_i) = 0$. The parameter $\gamma \in (0, 1]$ is acting as a filter, which slows down the growth of unbalanced value frequency distributions. With that, one can control the learning speed of the model with smaller γ leading to less penalization of less common values. In the experiment setup a standard value of $\gamma = 0.5$ is used.

3.4 Acceptance Strategy

Similarly to the case of opponent modelling, to maintain all other factors equal, the acceptance criteria are shared among all considered agents within this research paper. For the acceptance criteria, I have decided to implement a combination-based heuristic, as described in an article by T. Baarslag et al. [4]. Thus, agents always accept the bid when its utility is higher than the utility of a bid they are going to propose. Moreover, agents will always accept the final bid in the negotiation setting as the reservation utility equals 0, as mentioned in Section 3.1.

3.5 Eccentric Agents

In this section, I will elaborate on the specifics of the agents concerned only with their utilities. Thus, these agents will be used as a benchmark compared to the fairness-oriented agents to determine if the investigated metrics capture if the agents consider some fairness definition in their bidding strategy.

Simple Time-Dependent agent

The most straightforward agent that is only considering its utility, but at the same time makes some concessions to ensure an agreement, is the *Simple Time-Dependent agent*. It does not use the opponent modelling but makes the gradual concessions from its maximal utility by taking into account the elapsed time of the negotiations. The concessions are modelled using the time-concession function seen in the Equation 3.

$$T(t) = 1 - t^{\frac{1}{e}} \quad (3)$$

Equation 3: Function representing the level of concessions made depending on the elapsed negotiation time measured in elapsed negotiation rounds. Note that time t is normalized to fit range [0,1]

The time variable t increases linearly and mirrors the elapsed number of rounds, where the time $t = 0$ at round 0 and $t = 1$ at round 200. The variable e models the curvature of the made concessions. The closer the e is to 0, then at the beginning of the negotiations, the agent is less prone to make concessions. Such agents are called *Boulware* agents. On the opposite spectrum are the *Conceder* agents, which are characterized by the high value of $e >> 1$. When e is precisely 1 we can observe that the Equation 3 takes linear form of $T(t) = 1 - t$, which makes such agents referred in the literature as *Linear* agents. The different concessions curves are displayed in Figure 3 which compares different concessions strategies under various values of the variable e.

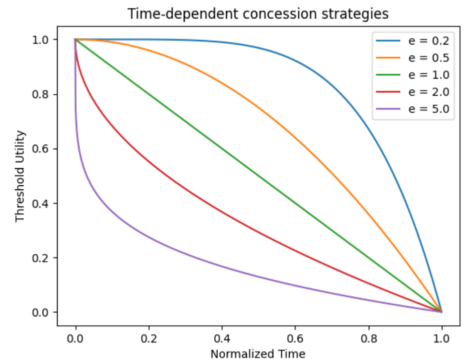

Figure 3: Various concessions strategies with respect to the value e as shown in Equation 3. The time-dependent agents with $e > 1$ are called Boulware Agents, and time-dependent agents with $e < 1$ are called Conceder Agents

In the experiment setup, I will use the Simple-time dependent agents with values of $e = 0.5, 1, 2$ and refer to them as Boulware, Linear and Conceder, respectively.

Hardliner agent

Hardliner Agent is a particular type of agent that, in principle, makes no or just minimal concessions from its own maximal possible utility value. Under the experiment setup, I use the GeniusWeb example Hardliner agent implemented with a Simple-time dependent agent with a low value of $e = 0.05$. With that, I expect the Hardliner agent to be very strict in his bidding strategy with a radical change in making concessions by the end of the negotiation. Although it may seem like an edge case behaviour, that agent's behaviour would be valuable to investigate with other agents as it would stress the agents and consider metrics under unusual situations.

Random Walker

The last eccentric agent considered in this experiment is the *Random Walker* agent, also included as a benchmark agent under the GeniusWeb framework. When making an offer to the opponent, it randomly chooses the first one that scores better than the utility value of 0.6 or takes the 20th considered bid if all previous fail to score utility higher than 0.6. With that, the negotiation setup supports an assessment of the fairness metrics if they can distinguish between fairness-oriented agents and the random negotiation trails generated by the Random Walker.

3.6 Fairness-Oriented Agents

This section considers the developed *Fairness-Oriented Agents*. The developed agents are much related to the optimal fair strategies found in literature or inspired by the agents from the annual, international Automated Negotiating Agents Competition (ANAC) [12] that show considerations of the fairness issues described by C. Albin as summarized in Section 2.1.

Non-Monotonic Concession Agent

The first considered fairness-oriented agent is an *Non-Monotonic Concession Agent* inspired by the HardHeaded agent from ANAC 2011 [22]. It is an agent that also uses concession function $T(t)$ as Simple Time-Dependent agent, but it switches between e_1 & e_2 values after some time t_0. The agent is designed to switch between concession strategies from Conceder to Boulware, so that $e_1 > e_2$. That change between concession strategies can be mathematically formulated as seen in Equation 4.

$$T(t) = \begin{cases} 1 - t^{\frac{1}{e_1}}, & \text{if } t \leq t_0 \\ (1 - T(t_0))(1 - (\frac{t-t_0}{1-t_0})^{\frac{1}{e_2}}) & \text{otherwise} \end{cases} \tag{4}$$

Equation 4: Function representing the level of the non-monotonic concessions made depending on the elapsed negotiation time measured in elapsed negotiation rounds. Once again note that time t is normalized to fit range [0,1]

Because of scaling the concession strategy after time t_0 by a factor of $(1 - T(t_0))$, I ensure the concession strategy is continuous. Thus, such scaling prevents the unexpected behaviour of the negotiating agents with a much-changing bidding strategy around time t_0. For a visualization of the non-monotonic concession strategies changing from Conceder to Boulware and vice versa, refer to the Figure 4.

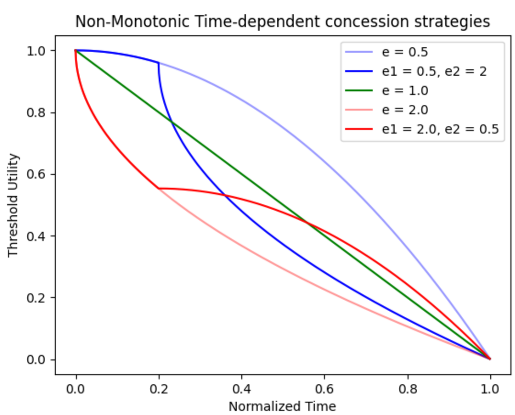

Figure 4: Comparison of the non-monotonic concessions strategies with their monotonic counterparts with respect to the value e_1 & e_2 as shown in Equation 4.

In this experiment setup we use the Non-Monotonic Concession Agent with $t_0 = 0.1, e1 = 1.2, e2 = 0.15$. By choosing such values of the variables, the agent allocates the first 10% of the negotiation time to present the bids that maximize their utility function so that the opponent could model its opponent's preferences accurately at the beginning of the negotiations. This strategy is an analogy to the considered Procedural Fairness issues. Additionally, agents try to give the bid close to modelled concession, but at the same time, that would be most beneficial for the opponent based on the Opponent Modelling.

Tit-for-Tat Agent

The next considered agent is the *Tit-for-Tat agent*, which focuses on Procedural Fairness issues and, in principle, reflects the strategy described by Baarslag et al. [3]. The basic procedure of the Tit-for-Tat agent's strategy is:

1. Measure the opponent's concession in terms of the agent's own utility function.
2. Mirror this bid by sacrificing the same amount as the opponent concedes.
3. Make the offer as attractive as possible for the opponent using opponent modelling.

The agent implements these principles as follows. If the agent starts the negotiation, it offers his maximal bid to the opponent. It also keeps track of the two last received bids to estimate the threshold of the offer to be made, as it tries to present an offer that mimics the concession made by the opponent. It does that by creating a small threshold window very close to the utility of the last offered bid minus the difference in own utility between the two last received offers made by the opponent. Then it selects all the bids that fit the threshold window. If no bids are in the threshold window, then the window size is multiplied by 2. However, if some bids fit the window, then it uses the opponent modelling to choose one of the bids that are most attractive to the opponent, which increases the chance of accepting an offer so that a fair consensus is achieved.

4 Results & Analysis

This section presents the results from the twenty negotiation tournaments in which every agent competed against every other considered agent with alternating utility preferences. The results presented in the tables have been sorted such that it firstly shows data for the eccentric agents and then the last two for fairness-oriented agents. They have also been marked with symbols E & F, respectively.

4.1 Distance to Bargaining Problem Solutions

Table 1, shown below, presents the results for the metric considering Distance to Bargaining Problem Solutions.

	Distance to Nash Point	Distance to Kalai-Smodrinsky
Hardliner (E)	0.257	0.302
Random Walker (E)	0.131	0.173
Boulware (E)	0.157	0.205
Linear Concsession (E)	0.142	0.205
Conceder (E)	0.155	0.200
Tit-For-Tat (F)	0.129	0.140
Non-monotonic (F)	0.137	0.183

Table 1: Aggregate results of Distance to Bargaining Problem Solutions

This metric consistently captures fairness in agents' behaviour, as fairness-oriented agents take top places of agents with the lowest distance to the solutions to the bargaining problem, apart from the Random Walker, which on average also obtains a low distance to the solutions of the bargaining problem. The filtering of random bids causes a Random Walker's high performance by thresholding low-utility bids; the bids have a high chance of being close to the Pareto frontier and thus close to the solutions to the Bargaining Problem. That is not likely to be a case for a pure Random Walker without limited bids filtering. However, on average, fairness-oriented agents obtain final allocations that are 21% closer to the bargaining problem solutions than those obtained from the eccentric agents, with the most radical difference being the distance to the Kalai-Smodrinsky solution.

4.2 Sum of Individual Utilities

Table 2 shows that the distinction between eccentric and fairness-oriented agents is not that straightforward for the sum of individual utilities. For example, simple time-dependent agents scored on average better than the considered fairness-oriented agents. A similar situation occurs when considering the distance to the point of maximal sum of individual utilities, with only a Random Walker being more off this point than the fairness-oriented agents. This suggests that considering only the Sum of Individual Utilities metric may not accurately capture fair behaviour for the negotiating agents.

	Sum of Utilities	Distance to Maximal Utilities Sum Point
Hardliner (E)	1.060	0.217
Random Walker (E)	1.070	0.250
Boulware (E)	1.080	0.165
Linear Concsession (E)	1.085	0.157
Conceder (E)	1.088	0.180
Tit-For-Tat (F)	1.070	0.240
Non-monotonic (F)	1.091	0.179

Table 2: Aggregate results of Mean Sum of Individual Utilities & Distance to Point of Maximal Sum of Individual Utilities

4.3 Time to Aggrement

The next metric is the mean time the agent needs to agree on a final bid with the results shown in Table 3. The time to agreement varies much across eccentric agents, whereas simple agents like Random Walker or Conceder are much more likely to settle early. In contrast, Hardliner takes the longest to find an agreement across all agents. The fairness-oriented metrics take high middle place in that category, making it hard to draw reliable conclusions about an agent's just behaviour based only on the needed time to reach an agreement.

	Mean Time to Aggrement
Hardliner (E)	194.2
Random Walker (E)	28.9
Boulware (E)	137.2
Linear Concsession (E)	86.6
Conceder (E)	45.1
Tit-For-Tat (E)	187.1‘
Non-monotonic (F)	151.7

Table 3: Aggregate results of Mean Time to Agreement for each participating agent

4.4 Kindness Function

The fairness-oriented agents should obtain higher values of the Kindness Function for constituting a reliable metric. However, as seen from Table 4, that is not entirely the case as Conceder and Random Walker obtain the highest results from all considered agents. When considering the Kindness function as a metric, keeping track of the Time to Agreement is beneficial to obtain a mean kindness per unit of one negotiation round. It also seems that mean kindness per round has a bit better performance than simple mean kindness per negotiation, with, in general, fairness-oriented agents scoring higher in kindness value than their eccentric counterparts. However, as Random Walker and Conceder also obtain higher values, the kindness function focuses on the proneness to find a compromise for a price of making concessions.

	Mean Kindness	Kindness per round
Hardliner (E)	-0.061	-0.00031
Random Walker (E)	0.307	0.01065
Boulware (E)	0.063	0.00046
Linear Concsession (E)	0.124	0.00143
Conceder (E)	0.170	0.00379
Tit-For-Tat (F)	0.292	0.00156
Non-monotonic (F)	0.129	0.00146

Table 4: Aggregate results considering kindness metrics both in mean simulation kindness and mean kindness per round

4.5 Fluctuations in final agreement

The last considered metric is the fluctuations in final agreements, with shown results in Table 5. The general trend is that the fairness-oriented agents are less volatile in their final settlement allocations. However, they obtained higher fluctuations than the Linear Concession agent or even Random Walker, which surprisingly obtained the lowest volatility score. The Linear Concession agent can explain these results as it tends to allocate equal contributions in utilities. On the other hand, the low fluctuations of a Random Walker are once again caused by the benchmark acceptance bid value above 0.6, which restricts the domain greatly from the inapplicable final allocations, and even if chosen by random, they are still considering a constrained set causing low standard deviation. Because of those surprising observations, one should refrain from relying only on fluctuations in the final agreement as a fairness metric and instead consider using it as a complementary fairness indicator.

	Standard Deviation in Utility 1	Standard Deviation in Utility 2
Hardliner (E)	0.179	0.154
Random Walker (E)	0.099	0.089
Boulware (E)	0.115	0.102
Linear Concsession (E)	0.099	0.094
Conceder (E)	0.124	0.115
Tit-For-Tat (F)	0.100	0.090
Non-monotonic (F)	0.120	0.112

Table 5: Aggregate results of the standard deviation of final utilities of Agent 1 & Agent 2

5 Discussion

Based on the results of the fairness metrics, it seems that the most reliable metric is the distance to the solutions to the bargaining problem, especially the distance to the Kalai-Smodrinsky point. However, as seen, it is also prone to some bias as it classifies the fairness of Random Walker as significantly high. That suggests that using discussed fairness metrics on their own can yield some errors. Thus, the main observation about the fairness metrics is that they should be considered in combination when assessing the fairness of an agent. For example, by combining the metric of distance to the solutions to the bargaining problem with the mean elapsed time to an agreement, the behaviour of Random Walker can be filtered by saying that an agent is fair if its final allocation is close to solutions to the bargaining problem and if it takes significant negotiation time to try on its opponent different bids. This observation is, however, too straightforward and may create errors when generalized to a more extensive set of considered agents. Thus, an exciting extension of this research would consider finding a model that takes into account interdependence between the fairness metrics. That could be organized by training a Machine Learning model, like a Neural Network, which would try to categorize agents as fairness-oriented and eccentric based on data from the fairness metric. Alternatively, one could create a regression model to assess fairness by taking multiple metrics as regressors. These extensions, however, are beyond the scope of this research. Additional improvements to the conducted research can consider the following ideas:

Repeat Experiment under different domains

Although the considered domain is extensive, with over ten million possible complete bids, and it takes the shape of an almost continuous convex set on utility space, a similar investigation could be performed under different domains. It may turn out that under an edge-case scenario, where there are no excellent compromise solutions (e.g. game-theoretic *Battle of the sexes* [6]), one could expect that there could be more fluctuations in the final utility allocations. For that, repeating the investigation under a different set of domains, which would vary in a number of possible complete bids and different utility space setups, would be beneficial to validate some of the findings described in this paper. Nevertheless, the domain for this research was carefully chosen to mirror most of the real-world negotiation problems.

Extend the number of considered agents

The research uses both eccentric agents as a benchmark for the fairness-oriented agents. When assessing the fairness metrics, it would be valuable to extend the number of considered agents, especially the fairness-oriented agents. By doing so, the findings made in the paper could be compared more reliably across the considered agents.

Allow agents to learn about opponents based on past encounters

In the investigated negotiations simulation scenario, we have treated each negotiation individually so that the negotiating agents could not store and later use the information about their opponent from the previous encounters. However, some agents could use that information to, depending on the opponents' strategy, offer bids that would be more beneficial to the opponent and more likely to accept given its negotiating strategy. That would require a change in the negotiating protocol to enable learning across negotiation sessions. This deviation could significantly change the experiment setup, which may impact the final results and observations based on them.

6 Responsible Research

There are no critical ethical issues concerning this research. Particular care has been put into correctly referencing all the ideas found in the academic resources. Additionally, all generated data comes from the negotiation simulations and has been included as a whole in the final calculation of fairness metrics. To allow reproduction of the research, the methodology has been thoroughly described in this paper, indicating the values for the used parameters describing the negotiation setup. Additionally, the reproducible code based on the *GeniusWeb* framework with considered negotiation agents and negotiation runner is in the GitHub repository (github.com/arubiobizcaino/FairnessInAutomatedNegotiation).

7 Conclusion

Due to many different ways of implementing the behaviour of the automated negotiation parties, there is uncertainty in estimating if the computational agents are behaving justly based on the individual fairness metrics. Nevertheless, the research has shown that the most consistent individual metric in fair behaviour estimation is the distance to the game-theoretic Bargaining Problem solutions (Nash Product, Kalai-Smodrinsky) related to the Outcome Fairness issues. Based on that, it would be the best fitted individual metric in automated negotiations to distinguish just behaviour in the participating agents. However, that metric has some bias as a Random Walker agent classified as an eccentric agent has, on average, scored close to these solutions. Thus, the accuracy in fair behaviour estimation can be improved by combining the fairness metrics like in the calculation of kindness function per negotiation round. However, considering all possible combinations of interdependence between metrics would require extensive research and could constitute a paper on its own. However, I hope that with this paper, the problem of lacking a generic framework for measuring fairness in automated negotiations has been tackled and that this paper could be used as a starting point for further research, especially with the growing shift in nowadays technologies to automated negotiations this topic would be fundamental and likely to be revisited.

References

[1] Cecilia Albin. The role of fairness in negotiation. *Negotiation Journal*, 9(3):223–244, 1993.

[2] Reyhan Aydoğan, David Festen, Koen V Hindriks, and Catholijn M Jonker. Alternating offers protocols for multilateral negotiation. In *Modern approaches to agent-based complex automated negotiation*, pages 153–167. Springer, 2017.

[3] Tim Baarslag, Koen Hindriks, and Catholijn Jonker. A tit for tat negotiation strategy for real-time bilateral negotiations. In *Complex automated negotiations: theories, models, and software competitions*, pages 229–233. Springer, 2013.

[4] Tim Baarslag, Koen Hindriks, and Catholijn Jonker. Effective acceptance conditions in real-time automated negotiation. *Decision Support Systems*, 60:68–77, 2014. Automated Negotiation Technologies and their Applications.

[5] Robert M. Coehoorn and Nicholas R. Jennings. Learning on opponent's preferences to make effective multi-issue negotiation trade-offs. In *Proceedings of the 6th International Conference on Electronic Commerce*, ICEC '04, page 59–68, New York, NY, USA, 2004. Association for Computing Machinery.

[6] Russell Cooper, Douglas V DeJong, Robert Forsythe, and Thomas W Ross. Communication in the battle of the sexes game: some experimental results. *The RAND Journal of Economics*, pages 568–587, 1989.

[7] Morton Deutsch. Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social issues*, 31(3):137–149, 1975.

[8] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, page 214–226, New York, NY, USA, 2012. Association for Computing Machinery.

[9] Jehoshua Eliashberg, Stephen A. LaTour, Arvind Rangaswamy, and Louis W. Stern. Assessing the predictive accuracy of two utility-based theories in a marketing channel negotiation context. *Journal of Marketing Research*, 23(2):101–110, 1986.

[10] Amitai Etzioni. Communitarianism. *Encyclopedia of community: From the village to the virtual world*, 1:224–228, 2003.

[11] Karen Huang, Joshua D. Greene, and Max Bazerman. Veil-of-ignorance reasoning favors the greater good. *Proceedings of the National Academy of Sciences*, 116(48):23989–23995, 2019.

[12] Catholijn Jonker, Reyhan Aydogan, Tim Baarslag, Katsuhide Fujita, Takayuki Ito, and Koen Hindriks. Automated negotiating agents competition (anac). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[13] Ehud Kalai and Meir Smorodinsky. Other solutions to nash's bargaining problem. *Econometrica: Journal of the Econometric Society*, pages 513–518, 1975.

[14] Johnathan Mell, Markus Beissinger, and Jonathan Gratch. An expert-model and machine learning hybrid approach to predicting human-agent negotiation outcomes in varied data. *Journal on Multimodal User Interfaces*, 15(2):215–227, 2021.

[15] John F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[16] Wouter Pasman, Pradeep K. Murukannaiah, and Catholijn M. Jonker. Geniusweb tutorial. *Interactive Intelligence, Delft University of Technology*, Sep 2021.

[17] Madan M Pillutla and J Keith Murnighan. Fairness in bargaining. *Social Justice Research*, 16(3):241–262, 2003.

[18] Matthew Rabin. Incorporating fairness into game theory and economics. *The American economic review*, pages 1281–1302, 1993.

[19] Victor Sanchez-Anguix, Okan Tunalı, Reyhan Aydoğan, and Vicente Julian. Can social agents efficiently perform in automated negotiation? *Applied Sciences*, 11(13):6022, 2021.

[20] Tuomas Sandholm. Automated negotiation. *Communications of the ACM*, 42(3):84–85, 1999.

[21] Okan Tunalı, Reyhan Aydoğan, and Victor Sanchez-Anguix. Rethinking frequency opponent modeling in automated negotiation. In Bo An, Ana Bazzan, João Leite, Serena Villata, and Leendert van der Torre, editors, *PRIMA 2017: Principles and Practice of Multi-Agent Systems*, pages 263–279, Cham, 2017. Springer International Publishing.

[22] Thijs van Krimpen, Daphne Looije, and Siamak Hajizadeh. *HardHeaded*, pages 223–227. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[23] Nancy A Welsh. Perceptions of fairness in negotiation. *Marq. L. Rev.*, 87:753, 2003.

[24] John Wilkerson. Rawls and nozick on fairness. *University of Washington*.

[25] Jonathan Wolff. Fairness, respect, and the egalitarian ethos. *Philosophy & public affairs*, 27(2):97–122, 1998.